Identification of strains of *Rhizobium leguminosarum* that are tolerant of desiccation

---

A thesis

submitted in partial fulfilment

of the requirements for the

Degree of Master of Science in Biochemistry

and Molecular Biology

at

Lincoln University

by

William Davies

Lincoln University

2018

Abstract of a thesis submitted in partial fulfilment of the requirements for the Degree of Master of Science in Biochemistry and Molecular Biology

Identification of strains of *Rhizobium leguminosarum* that are desiccation tolerant

By
William Davies

Agriculture in New Zealand relies heavily on mixed pastoral crops. Legumes, such as white clover in particular, are extremely important due to their symbiotic relationship with *Rhizobium* spp. Many of these pastoral areas can be subject to dry conditions and drought. For *Rhizobium* spp. it is important to understand what effect desiccating stress can have and whether there is natural variation in the ability of individual strains to tolerate such stress. This could inform the development of better commercial inoculants and enable them to persist better in dry soils. Numerous researchers have investigated the nodulation ability and nitrogen fixation efficacy of *R. leguminosarum* bv. *trifolii* when in symbiosis with *Trifolium repens* (white clover). However, little research has focused on the ability of *R. leguminosarum* to live in soil as a saprophyte. This study aimed to investigate variation in desiccation tolerance of individual strains of *R. leguminosarum bv. trifolli* and to determine if it was possible to identify strains with higher tolerance to this stress.

Twenty-four strains of *R. leguminosarum* were randomly selected from a collection obtained from sites that different in their annual soil moisture deficit (SMD). Twelve strains were chosen from three sites with high annual SMD (>100 days; dry sites) and 12 from three sites with low SMD (<5 days; wet sites). Sequencing of the 16S rRNA identified the strains as *R. leguminosarum*. Genotyping of the selected strains showed they were genetically diverse will 11 genetic groups identified, of which 6 were unique to a single strain. Two *in vitro* assays were used to identify strains that were tolerant of desiccation stress. The first assay measured relative biofilm (polysaccharide) production which has been linked to the ability to withstand desiccation stress in the literature. Following staining with crystal violet the results showed that strains produced significantly different quantities of biofilm (24 ($P \leq 0.001$), and 48 ($P \leq 0.001$) h incubation times) and this placed each strain into one of 8 different groups. The second assay investigated the ability of the strains to grow when placed under strong osmotic stress as osmotic stress involves a similar response pathway to desiccation stress in bacteria. The strains were incubated in two different concentrations of polyethylene glycol (PEG). There results showed there was a significant difference ($P \leq 0.005$) between the incubation times with more biofilm formed after 48 h than 24 h. There was also a significant difference in the amount of biofilm formed by individual isolates after 24 ($P \leq 0.001$), and 48 ($P \leq 0.001$) hours incubation, with the combined data placing the strains into 9 groups when grown in 50% PEG and 5 groups when grown in 60% PEG. The

strain that produced the most biofilm was 53 which originated from a wet site and the strains growing the most when exposed to PEG were 34 (50% Peg) and 42 (60% PEG).

To determine whether variation in genes producing the polysaccharide trehalose might explain variation in biofilm production or tolerance to PEG primers were developed to amplify five trehalose biosynthesis genes. These were trehalose-6-phosphate synthase (*otsA*), trehalose-6phosphate phosphatase (*otsB*), maltooligosyltrehalose synthase (*tre*Y), maltooligosyltrehalose trehalohydrolase (*tre*Z) and trehalose synthase (*tre*S). Only *ots*B produced readable DNA sequences suitable for analysis. When the DNA sequences were translated *in silico* there was a 98% identity between the amino acid sequences of the strains with six silent or conservative and two non-conservative substitutions. The two non-conservative substitutions were in isolates 32 and 42 and consisted of amino acid substitutions Serine to Alanine and Glutamic acid to Alanie. There was no obvious link between these amino acid substitutions and the ability of strains to produce more biofilm or grow in high concentrations of PEG.

To measure the survival of strains in soil spontaneous antibiotic mutants were created that were tolerant of erythromycin. The survival of the strains with the most biofilm formation (53) and greatest growth under the highest PEG concentration (42) were compared to two average strains and the commercial strain TA1. The biofilm assay was initially used to determine that the ranking of isolates and the relative ability of mutant and wild type did not differ. However, the results were inconclusive. The results showed that after 4 and 45 day in dry soil held at room temperature there were significant differences between isolates. After 4 d incubation strain 42 had a greater number of CFUs than isolate 47, and after 45 days strain 53 had a greater number of CFU compared to strain 50 and TA1. For all strains there was the expected decrease in CFU over time.

Overall this study identified that there is variation in the ability of strains of *R. leguminosarum* to withstand desiccation stress in soil. Strains 42 and 53 were selected using *in vitro* testing to be dry tolerant and survived better than other strains in dry soil. This indicated that the *in vitro* assays used here may be useful to rank strains of *R. leguminosarum* however, a greater number of strains, with more replication should be done to confirm this. This work highlights the potential of selecting strains to withstand desiccation and further development of a more robust screening process would assist desiccation tolerant commercial strains to be developed.

Keywords: *Rhizobium leguminosarum*, SMC, Biofilm, Polyethylene glycol (PEG), Polymerase chain reaction (PCR), trehalose-6-phosphate phosphatase (*ots*B), soil persistence

Acknowledgements

I would firstly like to acknowledge the Don Hulston Foundation for the financial support in my first year of my thesis.

I am grateful for the full support I received from my immediate family and close friends. Throughout the course of my studies made a big difference in motivating me to finish.

Thank you to the staff at Lincoln University who helped me during my time at Lincoln, especially Celine Blond and Sandy Hammond. Whose help, advice and encouragement were second to none.

Thank you to the other staff members that provided superb and efficient service, so I could get my results analyses quickly. I would like to also thank Candice Hume and Sandy Hammond for ordering lab supplies.

I would like to acknowledge my excellent supervisory team, Professor Derrick J. Moot, Dr Steve Wakeline for your guidance throughout my thesis. I would like to thank Assoc. Prof Hayley Ridgway for your expert guidance throughout the course of my thesis. I would not have finished without your positivity, enthusiasm and encouragement.

I would like to thank my bosses over the duration of my thesis. Your financial support made it possible to finish my thesis. A thank you to John Mckenzie for your patience whilst I finished my thesis

Table of Contents

List of Tables

List of Figures

Chapter 1: Literature review

1.1 Pastoral industry in New Zealand

New Zealand agriculture relies heavily on pastoral systems for stock production. It is essential to optimize the pasture growth. There are two farming systems that are typically implemented i) low intensity sheep, beef and deer (Matthews, Hodgson et al. 1999) farming (arid climate) ii) high intensity dairy farming (irrigated). Both of these systems implement mixed pasture, the most tradition mix pasture system is perennial ryegrass (*Lolium perenne* L.) and white clover (*Trifolium repens* L.). The white clover is added to supplement the most limiting factor, nitrogen. This system is not an efficient system due to limited access to water **(**Performance of subterranean and white clover varieties in dry hill country (Mills, Lucas et al. 2015).

Traditionally legume plants have been grown for the fruit or seed but within the last half century an emphasis has been put on a mixed pasture system. In New Zealand a significant increase of mixed pasture has been taking place for animal production. The legumes are being used in lowland dairy through to marginal dry land farming systems. There has been an emphasis of legume use in dry land farming due to difficulties of nitrogen applications. The use of legumes increases the soil nitrogen content via nitrogen fixation. There is also an ecological perspective, nitrogen leaching, and sustainability has also driven the use legumes in pasture opposed to chemical fertilizers (Zahran 1999, Wang, Yang et al. 2012).

1.2 Rhizobia

Rhizobia are gram negative, diazotrophic bacteria, belonging to the alpha-proteo family. They live as either saprophytes in the rhizosphere and rhizoplane (can be found outside rhizosphere), or as symbionts in the peribacteroid. Rhizobia can survive in soil (outside host rhizosphere) for several years (Downie 2010) but they do not produce spores or cysts to assist survival (Humann, Ziemkiewicz et al. 2009). They are motile when they are saprophytic, having active and passive movements (De Ley and Rassel 1965, Issa, Wood et al. 1993, Vlassak, Vanderleyden et al. 1997). Active movement is peritrichous movement in *Rhizobium spp.* (fast growing), and subpolar in *Bradyrhizobium spp.* (slow growing) (Vlassak, Vanderleyden et al. 1997). Otherwise they can move passively via water movement (Vlassak, Vanderleyden et al. 1997). They can move either randomly or via a chemotaxic direction (Vlassak, Vanderleyden et al. 1997). Rhizobia induce the formation of nodules when they form a symbiotic relationship with a specific legume species. Within the nodule of the specific species they develop a mutualistic relationship with the plant, in which they fix atmospheric nitrogen for the plant and the plant gives the bacteria nutrients and a protected niche to occupy. However, the rhizobia do not undergo vertical transmission. Instead they need to undergo a series of complex reactions to re-infect new leguminous plants. The reinfection can occur in two ways, crack entry or the more common way is nodulation.

The leguminous plants and rhizobia have co-evolved an important relationship. For this mutualistic relationship to be successful both plant and bacteria signal one another extensively. Microbes such as rhizobia target leguminous plants by producing a unique pathogen/ microbe associated molecular pattern (Hückelhoven 2007, Van Wees, Van der Ent et al. 2008, Faulkner and Robatzek 2012, Buscaill and Rivas 2014, Ökmen and Doehlemann 2014). This is of particular importance in agriculture because nitrogen is converted into a useable form for plants. This will furthermore boost the growth of the pastoral systems. Other legumes have been introduced into the pastoral systems such as Caucasian clover (*Trifolium ambiguum*), white clover (*Trifolium* repens) and lucerne (*Medicago sativa* L.), each with their own corresponding *R. leguminosarum* strain CC283b, *R. leguminosarum* strain TA1 and *S. meliloti* strain RRI128.

1.2.1 Nodulation

Low soil nitrogen concentrations cause the initiation of the symbiosis between legume and rhizobia. There are two ways in which rhizobia can infect, intercellular infection and intracellular infection (Suzaki and Kawaguchi 2014). Both methods of infection require infection and nodule organogenesis to occur before nitrogen fixation can occur.[1] Intercellular infections occur when the rhizobia enter through the apoplastic space of the outer epidermal cells and penetrate the inner cortical cells to form nodules (crack entry), however this infection process is poorly understood (Suzaki and Kawaguchi 2014). Intracellular (nodulation) infection is the more predominant and well-studied route. This procedure occurs as follows:

Legumes and rhizobia contain certain genes that encode products that act as chemical signals (Bardin, Dan et al. 1996). Legumes (legume root hairs) release specific chemo-attractants (Hirsch 1992, Cooper 2007) called flavonoids, which activate the rhizobia genes (Oldroyd, Harrison et al. 2005). The flavonoids stimulate a signal transduction cascade in the rhizobia that allows nodulation and nitrogen fixation to occur (Cooper 2007, Kobayashi and Broughton 2008). When the rhizobia detect the flavonoids, they undergo a form of quorum sensing to control population density (González and Marketon 2003, Downie 2010). When the bacteria are at enough density they activate five nod genes (Hirsch 1992, Stacey 2007, Maunoury, Kondorosi et al. 2008). The nod genes produce lipochitooligosaccharides (nod factors), each with a specific purpose. The nod factors contain specific chemical modifications between rhizobial strains, allowing host specificity (Hirsch 1992).

The nod factors are recognised via specific nod factor receptors located in the epidermal cells of the plant host. When the nod receptor recognises the nod factor morphological, electrophysiological changes occur within the root. There is a deformation of root hair tips, with isotropic growth where the root hairs have terminated growth. The calcium flux (Downie 2010) in

[1] **http://5e.plantphys.net/article.php?id=155**

the epidermis (electrophysiological change), alters gene expression. This results in depolarization of the plasma membranes and cortical cell division at nodule primordial (Cooper 2007). These processes allow the root hair to curl around the rhizobia population at the rhizoplane. The rhizobia are essentially trapped in an infection pocket. The specific rhizobia isolate may need to compete with other strains or species of nodulating bacteria (Downie 2010) for entry into the root hair. The rhizobia continue to replicate and pre-infection threads develop. The pre-infection thread causes localized cell wall degradation within the infection pocket which creates an invagination of the plant cell membrane (Sprent 2009).

The pre-infection thread also creates anticlinal orientated cytoplasmic bridges, which join the inner and outer sides of the cell. The pre-infection thread has an array of microtubules. The microtubules connect to the tip of the infection thread to the nucleus. The pre-infection thread acts as a guide for the infection thread allowing it to travel radially along the cytoplasmic bridges (Hirsch 1992, Maunoury, Kondorosi et al. 2008), toward the root cortex. The infection thread will elongate to the dividing root cortical cells. Rhizobia move along the infection thread towards cortex ((Krusell, Krause et al. 2005, Mergaert, Uchiumi et al. 2006). It is in this area, the inner cortex that the nodule develops (Maunoury, Kondorosi et al. 2008). The thread then branches out and produce un-walled infection droplets. These infection droplets contain rhizobia colonies. The rhizobia are incorporated into the cortical cells via endocytosis (Hirsch 1992, Oldroyd, Harrison et al. 2005, Sprent 2009), or more accurately endocytic budding (Downie 2010). The endocytosis process allows the plant cells plasma membrane to surround the bacteria colonies (Hirsch 1992, Sprent 2009). The bacteria multiple and the nodule is formed. This type of endocytosis creates a specialised membrane called the symbiosome membrane (Krusell, Krause et al. 2005). The bacteria multiple and the nodule is formed.

The peribacteroid becomes a specialised organelle similarly is the chloroplast (Simon-Rosin, Wood et al. 2003, Krusell, Krause et al. 2005). This specialised organelle contains two layers, plant plasma membrane and the bacteria's membrane (Udvardi and Day 1997). The space between is the peribacteroid space. Once this membrane is established nitrogen fixation can occur as illustrated by figure 1.1.

Figure 1.1 This is a schematic diagram of the symbiosome membrane. The plant cytosol is directly in contact with the peribacteroid membrane (symbiosome membrane). There is a space between the symbiosome membrane (peribacteroid space) and the rhizobia. The rhizobia is surrounded by a normal bacterial membrane (Udvardi and Day 1997).

The structure of this symbiosome allows the endocytobiotic association to occur (Udvardi and Day 1997) between the legume and rhizobia. However, since the nitrogen fixing process is very energy consuming (Vance and Heichel 1991), and the symbiotic continuum, tight regulation occurs between the two symbiotic partners. The onset of nodulation a variety of genes are activated, especially with rhizobia, such as the important nitrogenase gene (Bobik, Meilhoc et al. 2006), which are essential for nitrogen fixation. The plant also changes gene activity which helps maintain a mutualistic relationship. This helps prevents the rhizobia becoming parasitic. Proteins such as leghaemoglobins are believed to help maintain a healthy relationship by acting as an oxygen buffering protein. This is because nitrogenase is irreversible inhibited by oxygen (Rawsthorne, Minchin et al. 1980, Zahran 1999). The haemoglobin buffers the relationship as the prosthetic group has pentacoordinate iron binding (Hargrove, Brucker et al. 2000), which allows it to have a lower affinity for oxygen. The lower oxygen affinity means it can act as a buffering protein rather than an antioxidant. There are other proteins activated in both organism that helps maintain a healthy nutrient flow as illustrated via figure 1.2.

Figure 1.2 This is a schematic diagram of where the nutrients flow in a nodule. This diagram also illustrates what form the nutrients are transported as. The dashed lines represent minor steps. This diagram is from (Becana and Sprent 1987).

The symbiotic relationship between nitrogen fixing bacteria and legumes is an extremely complex interaction. It is not surprising that there are numerous factors have a significant influence towards the nodulation process. When the rhizobia are living as either saprophytes or endophytes they are continuously bombarded by interactions with other organisms, whether it is from similar bacterial strains or from parasitic microbes they must compete with for resources and space. These biotic factors are closely associated with the abiotic factors which also have a significant effect on the survival of the rhizobia in the soil and within the legume itself. These abiotic factors range from the nutrients available to the plant and or in the soil, temperature, and soil water content. A well-studied abiotic stress is desiccation, which maybe is a combination of several abiotic factors. Desiccation stress is one of the major factors affecting rhizobia (Osa-Afiana and Alexander 1982, Cytryn, Sangurdekar et al. 2007) populations. The abiotic and biotic factors can interact together to effect rhizobia survivability which will have a significant effect of overall legume and pastoral growth.

1.2.2 Biotic factors affecting rhizobia

The main effect of biotic factors is around competition either for survivability as a saprophyte or for ability to infect nodules. However, rhizobia are also preyed upon by some nematodes and

protozan species (Pena-Cabriales and Alexander 1979). Other species may have a beneficial on rhizobia such as Arbuscular mycorrhizal fungi that may help symbiosis between rhizobia and its specific host (Denison and Kiers 2011). It is suggested legumes influence specific rhizobia survival. This is suggested because rhizobia are not found in high concentration in the absence of host (Mary, Ochin et al. 1985). In the soil there are hundreds of microbes. These microbes are consistently competing for resources ranging from space to nutrients, especially in the nutrient dense rhizosphere. In order for rhizobia to be successful they need a broad range of metabolic pathways in order to utilize available nutrients in the rhizosphere. A broad range of metabolic abilities will also help rhizobia to adapt to environmental conditions, giving them a competitive advantage. Competition for water is a major factor.

The main form of competition is that for nodule occupancy. This competition could occur between or within species. This type of competition between species can occur due to the microbe associated molecular pattern. This means that all microbes emit a specific signal that allows the plant to recognise it as a beneficial or detrimental microbe. A pathogen either active or latent will try and mimic the signal of specific strains to enter legume plants. The competition can also be enhanced via horizontal gene transfer or more specifically transconjugation (Vlassak, Vanderleyden et al. 1997). This could involve the transfer of symbiosis genes into different bacteria (Downie 2010).

There is an alternative competition, same species competition. The successful nodulation and optimal nitrogen fixation occur with a specific strain of rhizobia for a specific legume cultivar. However, due to the high diversity of strains of *R. leguminosarum* in New Zealand soil (Howieson and Ballard 2004). This can influence the specificity at each stage of nodulation depending on intraspecific strains and their transconjugates (Martinez-Romero and Rosenblueth 1990). This strain variation can influence the rate of nitrogen fixation by the host.

The commercial strains of rhizobia have been selected to have a high nitrogen fixing ability. They have been used for this ability, but they have been shown to have low survivability in soil (Mary, Ochin et al. 1985). A better understanding of desiccation stress could help development of new commercial strains that are both desiccation tolerant and efficient at nodulation and nitrogen fixation.

1.2.3 Abiotic factors affecting rhizobia

There are numerous abiotic factors that are tightly intertwined that influence rhizobia survival. Some of the abiotic factors effecting rhizobia survival in New Zealand are soil nutrients, soil, water availability, and temperature. These factors can affect either the rhizobia saprophytic

survival or the nodulation process. Otherwise the abiotic factors could be limiting the symbiotic relationship, where the plant no longer requires nitrogen fixation.

Nutrients, especially macro nutrients, have a significant impact on the symbiotic relationship between rhizobia and legumes. The nutrients both parties are subjected to influence the interaction. For example, legume plants will not initiate the release of flavonoids if there is enough concentration of available nitrogen in the soil (Vlassak, Vanderleyden et al. 1997); this is of importance towards intensive agricultural systems such as dairy farms. Another example is the conditions for the rhizobia. Micronutrients effects on symbiosis i.e. iron, it has not been studied extensively (Geetha and Joshi 2013), but there are indications that it may influence rhizobia survival. Soils containing low iron may influence nodule formation and growth rate (O'Hara 2001, Geetha and Joshi 2013). Other nutrients; sulphur, phosphorus, calcium, boron, manganese (highly desiccation tolerant gram-positive bacteria accumulate manganese (OsaAfiana and Alexander 1982)) etc. (O'Hara 2001) have been shown to have an impact on the rate of nodulation and survival of rhizobia. Aluminium has been shown to have toxic effects for rhizobia (O'Hara 2001) and on some legumes. There is little evidence on the maximum levels of nutrients other than nitrogen and aluminium on rhizobia survival. This could be vital as less intensive agricultural systems; dry land systems may have toxic levels of micronutrients which could affect nodulation.

The soil can influence the survival of rhizobial communities when they are living as saprophytes (Bushby and Marshall 1977). Firstly, the soil type can influence rhizobia movement. This is because different soils have different water holding capacities. Although rhizobia contain a form of active transport (flagellation (Vlassak, Vanderleyden et al. 1997)), their movement is restricted via soil moisture content. The moisture content provides the bacteria with a form of movement but remain restricted in their microsite due to the rhizobia's discontinuous film (Vlassak, Vanderleyden et al. 1997). The movement can also be hindered by the charged soil colloids (Vlassak, Vanderleyden et al. 1997). Rhizobia can move between microsites via eukaryotic organisms such as earthworms (Vlassak, Vanderleyden et al. 1997). The soil pH can also affect the survival of rhizobia. The pH ties in directly with the nutrients. The availability of trace elements is dependent on soil pH. An example of this is aluminum. When the soil is acidic aluminum is readily available, which makes it toxic to rhizobia.

1.2.4 The effect of desiccation on rhizobia

The temperature (high temperatures) and water content combined are the most significant abiotic factor influencing the survival of rhizobia specially nodule establishment and growth (Vance and Gantt 1992)**.** Together these have a huge influence on plant and microbe survival. The combined water deficit and temperature cause dehydration of the bacteria. Desiccation stress is where the rhizobia are exposed to drying extremes for periods of time. The water deficit could be resulting from lack of water or extreme heat (including wind). Desiccation is the process of

dehydrating. The ability of the soil microbes to withstand this stress may not necessarily be drying out but the recovery to normal metabolism. This means that desiccation stress could also be considered as a cyclic stress of extreme drying and rewetting cycle (Potts 2001).

This is a significant stress for both legume plants and rhizobia in arid agricultural systems. Desiccation cause significant damage towards the bacteria and is one of the major factors affecting the legume-rhizobia symbiosis (McIntyre, Hore et al. 2007). The damage caused results in a decrease in the microbial efficiency (Steinweg, Dukes et al. 2013). Rhizobia require water as a hydration shell (Billi and Potts 2002). This hydration shell allows the structural integrity of the bacteria's membrane to remain intact. When the shell is removed the membrane is disrupted causing the van der Waal's interactions to occur (Potts 1999, Billi and Potts 2002). This interaction occurs between neighboring membrane lipids to alter their transition temperature (Billi and Potts 2002). This in turn causes a change (increase) in transition temperature of the membranes (Billi and Potts 2002). An increase in the transition temperature will cause the membranes to transition into a gel phase. The membranes that have transitioned into the gel phase will separate from those that have not. This gel phase will also lead to vesicle fusion (Potts 1994). This will lead to an aggregation of proteins. The bacteria undergo rehydration the membranes undergo another transitional phase which results in cell leakage (Billi and Potts 2002).

The reduced hydration leads to an increased number of free radicals in the cells. Increasing the free radicals could be formed via damaged membranes or altercations in protein structure (Billi and Potts 2002). The free radicals cause further damage to DNA and other proteins (Billi and Potts 2002). The damage to the DNA arises from cross-linking, and chemical modifications such as oxidation (Potts 1999, Billi and Potts 2002). The damaged DNA and proteins also mean the rhizobia are subjected to further damage from UV radiation (Billi and Potts 2002) resulting from gene mutation or damaged proteins.

Desiccation is a result of cyclic stresses involving a dehydration- rehydration process. The ability of bacteria to withstand desiccation is down to having a diverse array of physiological, morphological and diverse metabolism mechanisms to survive. The changes in function revolve around three phases, drying, storage and rewetting (Vriezen, De Bruijn et al. 2007, Humann, Ziemkiewicz et al. 2009). The drying stage of desiccation is of importance towards rhizobia survival. The survival is not only dependent on the bacteria stimulating a response, but survival is also dependant on how the rhizobia are dehydrated. The intensity of drying out can impact the survival (Vriezen, De Bruijn et al. 2007). This is reflective upon the time it takes to mount an effective response. The storage phase is where the rhizobia maintain a dormant metabolic state until there is enough moisture to maintain normal metabolic functions. The rewetting phase also has an impact on rhizobia survival. Like the drying phase the rate at which the rewetting occurs can affect survival. If rewetting is too fast the rhizobia may have disrupted subpolar regions

(Vriezen, De Bruijn et al. 2007). The same as the drying stage the slower the rhizobia are rehydrated the greater the survival. This process can be a cyclic process of continual drying and rewetting (Billi, Friedmann et al. 2001).

Each of these three phases invoke various response between species. For tolerance to desiccation the bacteria must be able to adapt to the following problems; accumulation of salts/ solutes, hyperosmotic stresses, impaired metabolism, and the accumulation of damaging molecules such as reactive oxygen species etc. (Vriezen, De Bruijn et al. 2007). Each of these issues require specific mechanisms to prevent the rhizobia from dying.

Rhizobia have a variation in their desiccation tolerance. This means that there are a variety of different responses to desiccation. Some responses are a production of antioxidants, clustering, change of nutrient metabolism and the production of disaccharides. These mechanisms help prevent damage within the bacteria and help it survive a period of dormancy.

1.3 Protection against desiccation stress

1.3.1 Production of antioxidants

The production of antioxidants helps prevents physical damage. These antioxidants can range from small molecules to proteins. Some bacteria (*Cyanobacteria* sp.) have been well characterised as being desiccation tolerant contain free radical scavengers that bind reactive oxygen species. The free radical scavengers may vary between and within species in accordance with the environment. Some bacteria may contain iron superoxide dismutase in certain soils (Vriezen, De Bruijn et al. 2007). Whereas in peat may contain a different scavenger such as the rhizobia's manganese superoxide dismutase (Vriezen, De Bruijn et al. 2007). In conjunction with this other protein maybe active, including UV absorbing pigment (Billi, Friedmann et al. 2001, Billi and Potts 2002).

1.3.2 Clustering

Another method for survival is clustering. Clustering or clumping is a mechanism some bacteria use as a response to stress. Another method for survival is clustering. Clustering or clumping is a mechanism some bacteria use as a response to stress (Liu, Gao et al. 2005). The mechanism is not well understood but may be a form of quorum sensing to allow the outer members of the cluster to sacrifice themselves so some of the individuals in the centre can survive. There may also be a change in the carbon source rhizobia use for their lowered metabolism during desiccation stress.

1.3.3 Polysaccharides

The excretion of saccharides has a significant effect on rhizobia survival when subjected to desiccation. The saccharides act as a carbon source, reduce redox damage, and to a certain

degree prevent dehydration. There are two disaccharides that have been characterized as having a significant impact on bacteria survival during desiccation, sucrose and trehalose. Trehalose has been associated with desiccation as having multiple functions associated with desiccation stress. This disaccharide is involved in forming glass formation (Cytryn, Sangurdekar et al. 2007), membrane stabilisation, preventing water loss, and acting as a carbon source (Argüelles 2000, Engelhard 2004, Reina-Bueno, Argandoña et al. 2012). It should be noted other disaccharides (sucrose) have been implemented in some of these processes as well.

Trehalose can form a glass transition state (Potts 1994). This transition state is also referred to as a biofilm. The trehalose's unique properties allow it to help bacteria survive desiccation. This biofilm is unique in comparison to other bacterial glasses because it has a much higher glass transition temperature (Potts 1994). The meaning of this glass transition temperature is similar to the membrane transition temperatures. Since trehalose has a much higher transition temperature in comparison to bacteria membranes, the membrane will not undergo a change into the gel phase, thus preventing vesical fusion (Potts 1994). This will in turn prevent membrane damage and cell leakage. The trehalose biofilm also helps prevent water loss. If the bacterial cell is dehydrated the trehalose can replace water as a hydration shell surrounding proteins etc. (Potts 1994). This is the water replacement hypothesis. The trehalose replaces the water surround molecules such as proteins and allows them to maintain function (Potts 1994). The hydration shell provided by trehalose also allows the metabolism to slow down (Potts 1994). Trehalose is essential for rhizobia to survive extended periods of desiccation. The trehalose hydration shell also prevents protein damage (Fernandez, Béthencourt et al. 2010), which will in turn lead to a decrease in free radical and UV damages.

This disaccharide is made up of two alpha glucose units linked by a specific 1-1 alpha glucoside bond (Argüelles 2000). Unlike other disaccharides it is a non-reducing sugar (McIntyre, Hore et al. 2007, Fernandez, Béthencourt et al. 2010), it has high hydrophilicity, and high stability (Argüelles 2000, Engelhard 2004). These properties mean it is of pivotal importance towards both prokaryotic and eukaryotic survive in extreme environments. It is also a potential explanation why there might be a range of desiccation tolerance levels.

1.3.4 Biosynthesis of trehalose

There are three biosynthetic pathways for trehalose production (Engelhard 2004, Cytryn, Sangurdekar et al. 2007, Reina-Bueno, Argandoña et al. 2012). The first pathway for trehalose is the OtsA/B pathway. This pathway involves the conversion of UDP-glucose into trehalose in a two-step process involving trehalose-6-phosphate synthase and trehalose-6-phosphate phosphatase enzymes (Engelhard 2004). This is the most common biosynthetic pathway. The other two biosynthetic pathways for trehalose found in bacteria are the MOTS pathway and the TreS pathway.

The MOTS pathway involves an intermolecular transglycosylation process (Engelhard 2004). This involves the conversion of a maltooligosyltrehalose (larger than four-unit polymer) into trehalose via maltooligosyltrehalose synthase and MOT trehalohydrolase enzymes. The TreS pathway is similar to the MOTS pathway as it uses an intermolecular transglycosylation of a maltose unit, using trehalose synthase (De Smet, Weston et al. 2000, Engelhard 2004).

Figure 1.3 This is a schematic diagram showing the 3 biosynthesis pathways for trehalose in bacteria. Pathway a) is the most common being found in numerous microbes. Pathway b) is found in a hand full of microbes (including rhizobia). Pathway c) has been found in Pimelobacter sp. and in Thermus sp. (De Smet, Weston et al. 2000).

Since trehalose is abundant throughout the terrestrial life forms, has numerous functions in desiccation and has several biosynthesis pathways make it an ideal candidate to help determine and understand desiccation tolerance between rhizobia strains. It is also an important aspect to study for desiccation tolerance in rhizobia because they can synthesize trehalose both as a saprophyte and a symbiont (Suárez, Wong et al. 2008). This has huge implications, especially if the desiccation susceptible commercial strains can be manipulated in such a way to improve tolerance.

1.4 Methods to identify desiccation tolerance in microorganisms

It is important to be able to identify isolates being used throughout the experiment when looking at a response to a stimulus. One way of doing this is to obtain a phenotype. This thesis aims to characterize isolates that are desiccation tolerant by establishing a phenotype and methodology to indicate tolerance.

Rhizobia are a diverse family of bacteria containing over 120 species (O'Hara, Zilli et al. 2016), found with 15 genera found in seven families (Giller, Herridge et al. 2016)**.** This first step is to identify the species via housekeeping genes (Gaunt, Turner et al. 2001). The use of the polymerase chain reaction (PCR) has revolutionized our ability to study variation in rhizobia genetics. It has allowed individual isolates of rhizobia to be retrieved from soil and nodules, and then identified (Hebb, Richardson et al. 1998). Examples to PCR use are 16s rRNA (Jensen, Webster et al. 1993) which identifies species via targeting housekeeping genes, and enterobacterial repetitive intergenic consensus (ERIC) (Versalovic, Schneider et al. 1994). ERIC PCR allows short repeats to be identified. This allows the rhizobia to be categorized into phenotypes. Genes can also be targeted to further characterize variation between strains.

In conjunction with DNA sequencing of PCR fragments can be done. This has revolutionized the analysis of microbial communities (Zepeda Mendoza, Sicheritz-Ponten et al. 2015). The use of next generation sequencing allows multiple DNA sequences millions of nucleotides long to be analyzed with one another (Shokralla, Spall et al. 2012)**.** This is of particular importance when analyzing variation in genes.

It is difficult to analyse desiccation tolerance in soil due to the numerous variables that can influence the results (Deaker, Roughley et al. 2004). Rhizobia are sensitive to environmental stress and die easily in unfavourable conditions (O'Hara 2001). It is important to be able to do rapid screening *in vitro* to give an indication of desiccation tolerance. One way to mimic water stress by manipulating osmotic potential (Blum and Ebercon 1981, Bayoumi, Eid et al. 2008) between the bacteria and the environment it lives in using Polyethylene gycol (PEG).

Another way to characterise tolerance is to measure a response to desiccation. The formation of extra cellular biofilms has been reported to have an influence on desiccation tolerance (Vanderlinde, Harrison et al. 2010). Measuring the biofilm can indicate desiccation tolerance (Sandhya, Grover et al. 2009).

1.5 Rationale of this research

The use of legumes has inherent limitations in marginal dry lands. The high country has restricted water access. Thus, the rhizobia within the legume may not be present or able to survive the unforgiving dry land environment. By identifying rhizobia tolerant of dry soil, they may be able to persist better in marginal soils. This will improve the fertility of soil and increase animal production in a sustainable manner. The overall goal of this research is to identify desiccation tolerant isolates of *Rhizobium leguminosarum*. To achieve this thesis has three main objectives:

1. To develop bioassays for traits associated with desiccation tolerance by strains of *R. leguminosarum*.
2. To sequence candidate genes involved in the production of trehalose by *R. leguminosarum* and to determine whether polymorphisms are associated with tolerance to desiccation
3. To investigate whether strains of *R. leguminosarum* selected for their ability to withstand desiccation persist longer in dry soil.

Chapter 2 - Bioassay of strains of *Rhizobium leguminosarum* for traits associated with desiccation tolerance

2.1 Introduction

Soil dwelling microbes are consistently subjected to a vast array of stresses. An example a one of these stresses is desiccation. Desiccation stress invokes a complex range of biochemical responses to help survive decreased water potential. These responses may focus on repair systems or help maintain a state of metabolic dormancy (Potts 1994). Recent attention has been paid towards the ability of prokaryotes to produce extracellular polysaccharides as a mechanism of protection against desiccation. In particular trehalose has been indicated as having a possible role in desiccation tolerance (Cytryn, Sangurdekar et al. 2007). One of the roles of trehalose in desiccation tolerance is the production of an extracellular biofilm. It is also acknowledged that many factors can help determine tolerance to dry conditions such temperature and oxygen concentrations in soil (Vriezen, De Bruijn et al. 2007).

When rhizobia are subjected to desiccation their external environment has a significant decrease in water potential. An increase in the salinity in which the rhizobia are grown in can simulate a reduction in water potential and therefore help simulate desiccation stress (Radhouane 2007). Survival to external salinity conditions can indicate desiccation tolerance. Polyethylene glycol can replicate very high osmotic pressures (Van den Berg and Zeng 2006). The presence of root nodules can also give a relative indication of rhizobia desiccation tolerance. If saprophytic rhizobia are left in dry soil for a period and they have a low tolerance to desiccation they will be in lower numbers in the soil. If they have a low concentration in the soil, there should be fewer nodules. This thesis is focused on identifying rhizobia strains that differ in their tolerance to desiccation within a culture collection of 520 strains that were sourced based on annual rainfall data. In the collection of 520 strains there were strains from 26 sites that ranged in soil moisture deficits (SMD) from 149 days per year (very dry) to 0 (wet) (van Ham, O'Callaghan et al. 2016). A general assumption can be made with reference to the sites that the rhizobia will have experienced different selection pressure which will generate genetic diversity. Within this genetic diversity there will be a range of strains that are desiccation tolerant and intolerant.

The aim of this chapter was to test genetically different strains selected from the culture collection of rhizobia for their desiccation tolerance. To characterize the isolates into desiccation tolerant and intolerant strains two experiments were a conducted i) Ability of strains to form a biofilm and ii) Ability to grow in the presence of high concentrations of polyethylene glycol. The overarching hypothesis is that *"strains of* R. leguminosarum *originating from areas that receive low annual rainfall will show evidence of desiccation adaptation."*

2.2 Materials and Methods

2.2.1 Selection of *Rhizobium leguminosarum* strains

Four collection sites were chosen in accordance with their annual soil moisture (two sites with low moisture and two sites with high moisture) (van Ham, O'Callaghan et al. 2016). A total of 24 strains of *Rhizobium leguminosarum* were selected randomly from the four chosen collection sites (6 isolates per site), using a random number generator. The 24 strains were originally isolated from white clover plants grown in soil collected from the four sites.

Each isolate had been stored in 20% glycerol at -80°C. Each strain was sub-cultured by streaking a loop of bacterial culture from the -80°C stock onto an Petri plate containing sterilised yeast mannitol agar (YMA: 10 g mannitol, 1 g yeast extract, 0.5 g dipotassium phosphate (K_2HPO_4), 0.2 g magnesium sulphate ($MgSO_4$), 0.1 g sodium chloride (NaCl), 1 g calcium carbonate ($CaCO_3$), 15 g agar (standard Davis agar), and 1 L water). The details of each site are given in table 2.1.

Table 2.1 Details of the sites from which the *Rhizobium leguminosarum* strains were sourced from this study.

Location	Isolate numbers	Soil Moisture Deficit (days)	Most Probable Number* (MPN)
Dry Sites			
Picton-Awatere Valley (6)	31-36	120-140	2.27×10^4
Central Otago-Middlemarch (17)	37-42	100-120	5.77×10^4
Wet sites			
West Coast-Whataroa (12)	43-48	<5	3.74×10^7
New Plymouth-Mildhurst (26)	49-54	<5	7.31×10^4

* Estimated number of *R. leguminosarum* per g soil

2.2.2 Identification and genotyping of the selected strains

2.2.2.1 DNA extraction

A single colony from a 2-3-day old culture of each strain grown on YMA was placed in 1 mL of yeast mannitol broth (YMB) in a sterile 1.7 mL tube. The tube was shaken at 220 rpm (LABNET 211 DS, Labnet International, USA) at 28°C for 24-36 h. A control tube containing only YMB was included to check for any contamination. DNA was extracted from broth cultures using the PUREGENE DNA extraction kit (Qiagen) according to the manufacturer's instructions. Briefly, the incubated broth culture was centrifuged at 13,500 x g for 2 min and the supernatant discarded. To the pellet of cells 300 µL of Cell Lysis solution was added and pipetted to mix. The tubes were incubated for 5 min at 70°C and add 1 µL of RNaseA solution (10 mg/mL; Invitrogen) was added to the cell lysate solution. The tube was inverted 25 times and incubated at 37°C for 15 min. After cooling to room temperature, 100 µL of Protein Precipitation solution was added and vortexed for 15 s. The samples were centrifuged at 13,500 x g for 3 min to pellet the proteins and cell

debris. The supernatant was transferred to a sterile 1.7 mL tube and 300 µL of ice cold 100% isopropanol added to precipitate the DNA. The tube was gently inverted 50 times and then centrifuged at 13,500 x g for 1 min to pellet the DNA. The supernatant was discarded and 500 µL of 70% ethanol added. The tubes were centrifuged again at 13,500 x g for 1 min. The supernatant was discarded, and the pellet air-dried before resuspending in 30 µL of sterile water prior to storing at 4°C. The DNA concentration was measured using a Nanodrop spectrophotometer (NanoDrop Technologies Inc., Delaware, USA) and aliquots adjusted to 10 ng/ µL (16S PCR) and 50 ng/µL (ERIC-PCR).

2.2.2.2 16S rRNA PCR

The 16S rRNA gene region was amplified using primers F27 (5'-AGAGTTTGATCMTGGCTCAG-3') and R1494 (5'-CTACGGYTACCTTGTTACGAC) (Weisburg, Barns et al. 1991). Each 25 µL reaction contained 1 µL of each primer (10 µM), 2 µL of 2.5 mM dNTP's, 2.5 µL 10x buffer (FastStart, Roche, USA), 0.25 µL of Taq polymerase (5U/µL; FastStart, Roche, USA), 17.25 µL of sterile H_2O and 1 µL of sample DNA (10 ng). A negative control in which 1 µL of sterile H_2O instead of DNA was added was included to check for contamination.

The tubes were placed in the thermal cycler (Applied Biosystems Veriti 96-wells thermal cycler) and the programme, 94°C for 3 min, followed 38 cycles of 94°C for 30 s, 55°C for 30 s and 72°C for 1 min and a final 72°C for 7 min was run. The PCR products were stored at 4°C until gel electrophoresis as described in section 1.2.2.4.

2.2.2.3 ERIC-PCR

A genotype for each isolate was generated by ERIC-PCR to allow isolate verification in future work. The DNA was amplified using primers ERIC 1R (5'-ATGTAAGCTCCT GGGGATTCAC-3') and ERIC 2 (5'-AAGTAAGTGACTGGGGTGAGCG-3') as described by (Versalovic, Koeuth et al. 1991). Each 25 µL reaction contained 1 µL of each primer (50 µM), 2 µL of 2.5 mM dNTP's, 2.5 µL 10x buffer (FastStart, Roche, USA), 0.25 µL of Taq polymerase (5U/µL; Roche), 17.25 µL of sterile H_2O and 1 µL of DNA (50 ng/µL). A negative control was included to check for contamination. The tubes were placed into a thermal cycler (Applied Biosystems Veriti 96-wells thermal cycler) and the programme was 95°C for 3 min then 40 cycles of 95°C for 30 s, 52°C for 30 s, 72°C for 1 min and a final 72°C for 10 min was run. The PCR products were stored at 4°C until gel electrophoresis as described in section 2.2.2.4.

2.2.2.4 Gel electrophoresis

The PCR products were run on a 1% agarose gel Progen Biosciences, Brisbane, Australia). This gel was made by dissolving 1 g agarose in 100 mL 1xTris-acetate-EDTA buffer (1x TAE; 40 mM Tris acetate, 2 mM Na_2EDTA, pH 8.5). Six □L of each PCR product (including control) was added to 2 □L

of loading dye, mixed by pipetting and loaded into individual wells on the agarose gel submerged in 1x TAE buffer. The 1 KB plus DNA ladder (Invitrogen) was prepared in the same manner and loaded into one well per gel for size comparison. The gel was run at 10 V/cm for 55 min. The gel was stained with 0.5 µg/ mL ethidum bromide for 15 min, rinsed in tap water and photographed under UV light using a Firereader (UVITEC Gel Documentation Systems).

2.2.2.5 Sequencing

The 16S PCR products were sequenced directly by the Bio-Protection Research Centre Sequencing Facility, and Lincoln University in both direction using primers F27 and R1494. The resultant sequences were edited using DNAMAN Version (Lynnon Biosoft version 4) and Chromas lite (Technelysium Pty Ltd) to remove any poor or ambiguous sequence. The DNA sequences were compared to other known sequences in GenBank using the BLASTn algorithm on NCBI (http://blast.ncbi.nlm.nih.gov/).

2.2.3 Bioassays to rank strains for desiccation tolerance

Two methods were used to test the ability of each strain to withstand desiccation. The combined data was used to group isolates into dry tolerant and intolerant strains.

2.2.3.1 Biofilm formation by strains of *R. leguminosarum*

2.2.3.1.1 Culture preparation

Each bacterium was inoculated into 1 mL of YMB in sterile 1.7 mL tubes and grown for 48 h at 28°C in an oscillating incubator (110 rpm) to reach stationary phase (OD_{600}≥1.0). The absorbance of each culture was measured at 600 nm in a spectrophotometer. The bacterial broth cultures were adjusted to the same concentration (1× 10^8 CFU/mL) using YMB. Each culture was the diluted 1:100 using fresh assay medium (YMB) to a concentration of 1×10^6 CFU/mL. A 100 µL aliquot of diluted culture was placed into a 96-well microtiter plate (brand) and replicated on each of 6 microtiter plates. Six corresponding wells containing uninoculated YMB medium were used as a control. Each plate was covered with a classic seal plate sealing films and incubated at 20°C in the dark. Duplicate plates were measured after 24, 48 or 72 h incubation.

2.2.3.1.2 Biofilm staining

After incubation cells and excess YMB were removed by inverting the 96-well plate and shaking gently by hand. The 96-well plate was rinsed by submerging in sterile water. While the plate was submerged the surface of the plate was gently rubbed with gloved fingers to release bubbles and ensure water entered all wells. After rubbing the plate was removed from the water, inverted, and excess water removed by gentle shaking. This process is repeated once so that plate is rinsed two times. The plate is then turned face down and patted firmly on a paper towel to remove any

remaining water. To each well 125 µL of 0.1% crystal violet (McDougall, 2015) solution was added and incubated at room temperature for 10 min adapted from Coffey and Anderson (2014). The plate was then inverted and shaken gently to remove liquid. The plate was then rinsed in water thrice as previously described before being placed face down on a paper towel to remove excess water. Each 96-well plate was dried for 24, 48 and 72 hours at room temperature.

2.2.3.1.3 Optical density measurement

After drying for 24 hours 150 µL of 30% (v/v) glacial acetic acid was added to each well, pipette mixed and left to sit for 10-15 min. Then 125 µL of each sample was transferred to a flat bottom, optically clear 96-well micotiter plate and the absorbance measured at 590 nm and/or 550 nm. The same process was repeated for the 48- and 72-hour incubation times

2.2.3.2 Tolerance to Polyethylene glycol (PEG)

2.2.3.2.1 Solution and bacteria preparation

Polyethylene glycol 6000 (PEG) mixtures were made to 50 and 60% (w/v) in YMB and autoclaved to sterilise. A single bacterial colony of each strain were inoculated into 1 mL YMB in sterile tubes and grow to stationary phase which was typically 24 to 36 h at 28°C and 120 rpm. The bacteria were diluted to 1×10^7 CFU/mL.

2.2.3.2.2 Incubation

For each strain a duplicates tubes were prepared each containing a 100 µL aliquot of 1×10^7 CFU/mL added to 900 µL of 50 or 60% PEG solution in a 2 mL tube and placed at 28°C and 180 rpm (brand) for 4, 8 14, 24 and 48 h respectively. At each time point the optical density was measured using a spectrophotometer at 600nm.

2.2.3.2.3 Statistical analysis

Statistical analysis was done using the GenStat programme. A two-way ANOVA was used to analyse the variation between treatments using a Fisher's Protected Least Significant Difference (5% Least Significant Difference value). The isolates and incubation times are the fixed factors, whereas the absorbance is the variant.

2.3 Results

2.3.1 Genetic diversity

2.3.1.1 Identification of bacterial isolates by DNA sequencing

All the strains produced a single PCR product of approximately 1400 base pairs as illustrated by figure 4. For each of the strains a 1400 base pair PCR product was sequenced. The usable sequences produced were assembled. A comparison with sequences on GenBank showed that most strains were *Rhizobium leguminosarum* (99-100% identity) with the exception of isolates 38, 41, and 47 (GenBank identity >90%). Isolate 52 GeneBank search indicated is was *Massilia* sp. (GeneBank identidy 98% Accession number EU420063.1). Sequences were 91% similar to each other when aligned with isolate 52. There was a 98.15% similarity between sequences when isolate 52 was removed. There was no grouping between dry and wet sites.

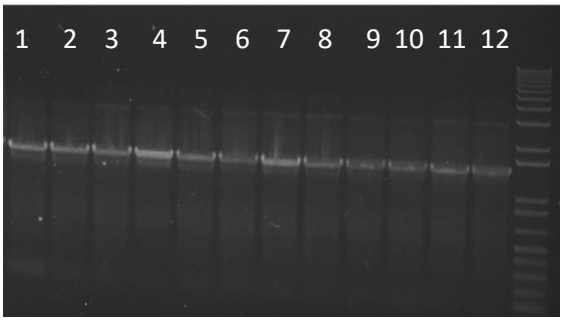


Figure 2.1 Agarose gel (1%) of the 16S PCR of 12 strains of *R. leguminosarum*. Lanes 1 through 12 are isolates being used; Lane 13 is the 1 KB plus DNA ladder (Invitrogen).

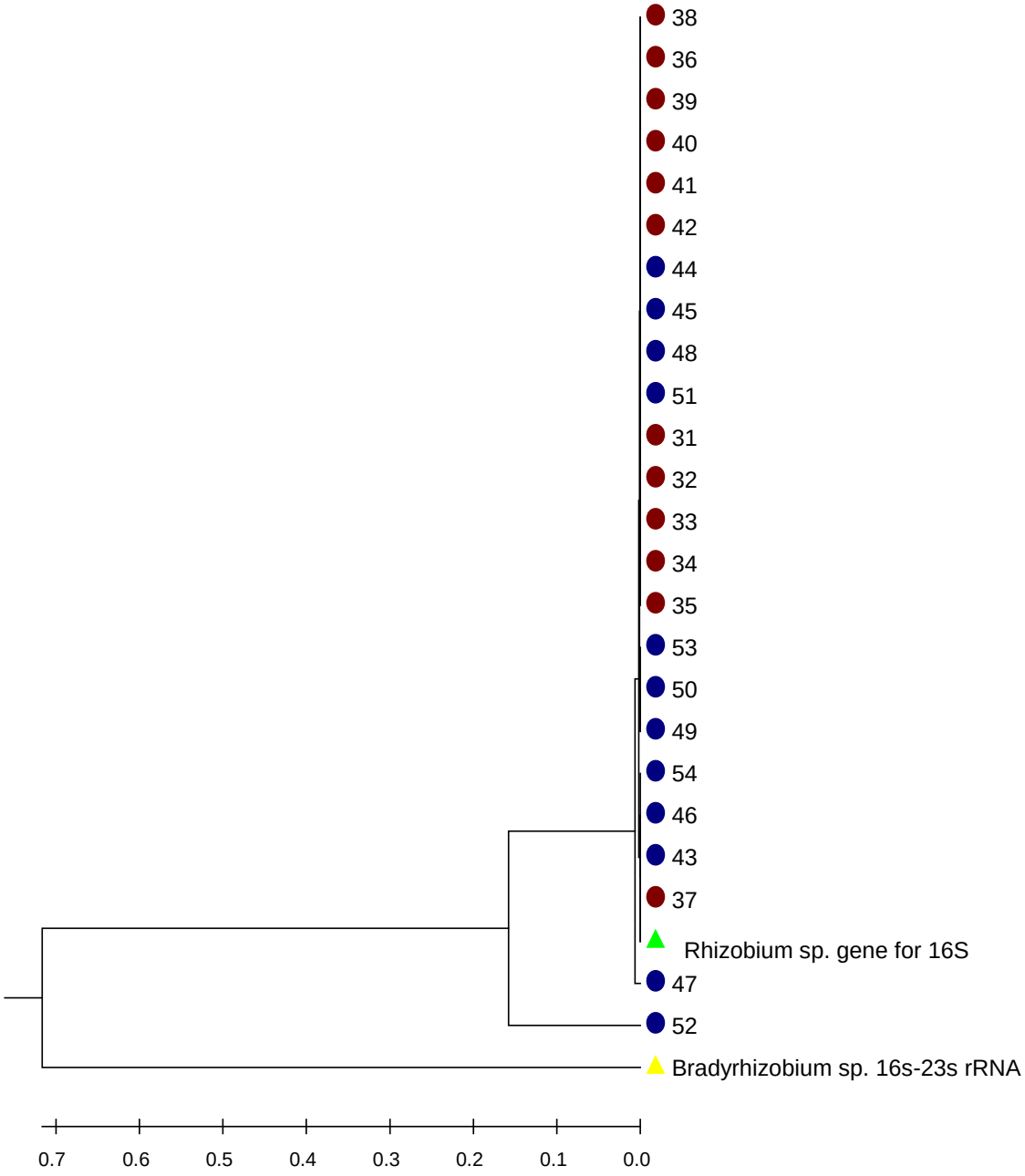


Figure 2.2 Phylogeny tree from 16s sequence data. Outgroup (yellow circle) consisted of Uncultured *Bradyrhizobium* sp. clone WME3 16S ribosomal partial gene sequence (Accession number FJ544532.1). Type strain (green triangle) consisted of *Rhizobium leguminosarum* bv. trifolii clone (Accession Rt621KX486575.1). Wet site isolates (blue circle), dry site (red circle).

2.3.1.2 Genotyping of isolates

All the isolates produced an ERIC-PCR fingerprint. The isolates were placed into 11 fingerprint groups and these were group A (31), group B (32, 35, 36), group C (33) , group D (43,48,52), group E (39, 40, 45, 46, 47, 49, 50, 51), group F (34, 37), group G (44, 53), group H (38), group I (41), group J (42) and group K (54). Group E was the largest containing 8 of the 24 isolates and six isolates were unique. Each site had between 3 and 5 genetically distinct isolates, with site 17 being the most diverse with 5 different genotypes present. These genetic fingerprints did not correlate with the site they isolates were collected from.

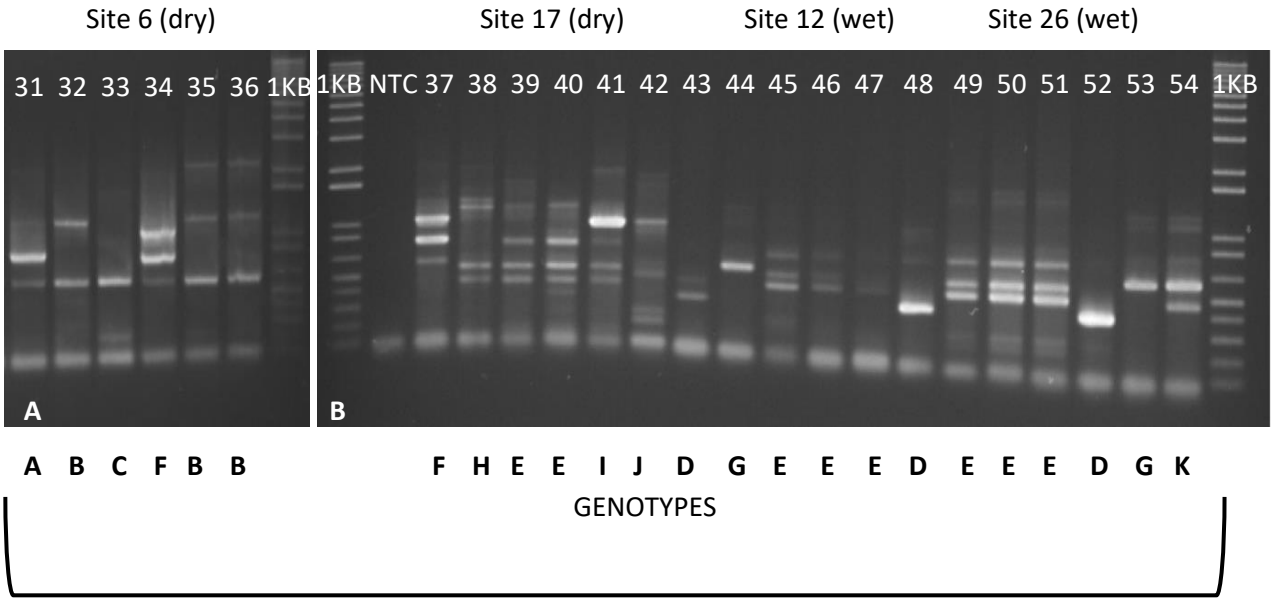


Figure 2.3 Agarose gel (1%) of DNA from strains of *Rhizobium leguminosarum* amplified by ERIC-PCR. Gel A consists of isolates 31-36 (from site 6) plus a 1kB plus DNA ladder. Gel B consists isolates 37-42 (site 17), isolates 43 - 48 (site 12), and isolates 49 - 54 (site 26) plus two lanes of 1kB plus DNA ladder, and a negative control (NTC). Genotype group is designated at the base of the gel.

2.3.2 Desiccation Tolerance Ranking

2.3.2.1 Formation of a polysaccharide biofilm by strains of *R. leguminosarum*

Prior to spectrophotometry a visual assessment of the stained biofilms was done. All the isolates produced biofilm as evidence by the purple stain produced by the crystal violet. The visual assessment of the isolates when incubated for 24 h incubation provided a presence or absence as many of the films were not robust enough for the washing process. The negative control also showed some mild staining. After 48 hours incubation visible stains were present for all the isolates. The isolates (including the negative control) could be placed into 2 groups according to the stain's colour intensity. The highly stained (strong colour) group consisted of isolates 31, 33, 34, 35, 37, 38, 39, 41, 42, 47, 48, 49, 51, 52, 53, and 54. The poorly stained group (pale colour) consisted of isolates 32, 36, 40, 43, 44, 45, 46, and 50 (also the negative control). There was a

significant difference (P≤0.005) between the incubation times with more biofilm formed after 48 h than 24 h. There was significant difference in the amount of biofilm formed by individual isolates after 24 (P≤0.001), and 48 (P≤0.001) hours incubation (Table 2.2). The significant differences in absorbance place the isolates into 7-8 groups. Group 1 represented isolates that produces the least amount of biofilm and increased in ascending order to group 7/8 which contained isolates that produced the most biofilm. The absorbance after 24 hours incubation time for the smallest biofilm was 0.2305 - 0.2567 and 0.3097 for the most biofilm (appendix 7.2.2). The absorbance after 48 hours incubation time for the smallest biofilm was 0.2636 -0.3344 and 0.504 - 0.5046 for the most biofilm (appendix 7.2.2). The average absorbance for isolates in group 1 after 24- or 48-hours incubation was ≤ the control.

Table 2.2 Statistically significant groups of isolates for average biofilm formation after 24 h, 48 h and from combined data. Groups are in ascending order from least biofilm produced (group 1) to most biofilm produced (group 7/8). Isolates in red have originated from dry sites and isolates in blue from wet sites.

Group	Incubation time		
	24 h	48 h	Combined (24 + 48 h)
1	36, 46, 47	34, 36, 37, 38, 39, 40, 41, 43, 44, 45, 46, 47, 48, 49, 50, 51	32, 36, 38, 46,39, 44, 43,46 , 47, 48, 50
2	31, 32, 33, 39, 41, 43, 44,	32, 42	31, 37, 40, 45
3	34, 37, 38, 49, 50, 54	31	34
4	35, 51	52	41, 49, 51
5	45, 52	35	33
6	40	33	35, 42
7	53	53, 54	52
8	42		53, 54

When the data from 24- and 48-hour incubation times was combined there were 8 biofilm groups produced (Table 2.2, Figure 2.4). There was a significant difference between sites (P≤0.001). The sites were placed into 3 groups with isolates from site 12 (wet site) producing the least biofilm, the middle group consisting of isolates from site 17 and site 6 (both dry sites) and the most biofilm produced by isolates from site 26 (wet site).

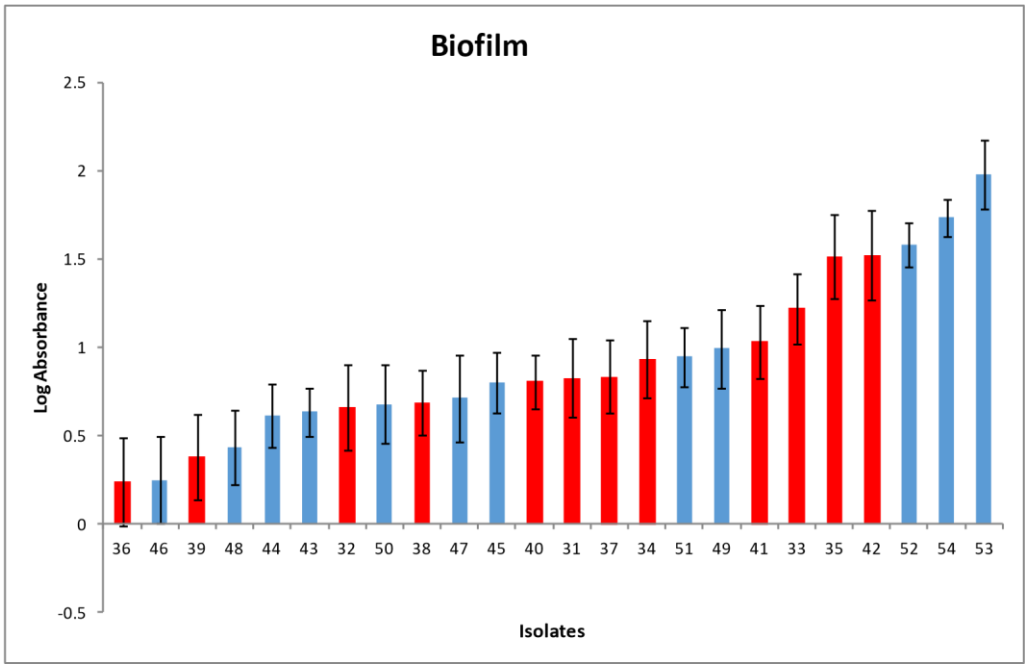


Figure 2.4 Relative biofilm produced when Rhizobium leguminosarum sp. incubation times were combined, and time was included as a statistical factor. The control absorbance has been deducted from the isolate absorbance. Samples were divided into either dry site collection (red bars) or wet site collection (blue bars).

1.3.2.2 Polyethylene glycol

There were significant differences in the ability of different strains to grow in the 50% and 60% PEG solutions. The 50% PEG concentration showed no statistical significance (P≥0.05) at the 4 hour and 24-hour incubation times. Significant differences were observed after 6 hours (P≤0.001), 14 hours (P≤0.001) and 48 hours (P≤0.001) incubation. The isolates were placed into different statistically significant groups for each incubation time (Table 2.3). Growth in 50% PEG for 6 hours produced 10 groups, for 14 hours incubation produced 7 groups and for 48 hours produced 9 groups. The relative growth between strains over time is shown in figure 2.4. When strains were grown in 60% PEG there were no significant differences between strains grown in the 24 hours (P≥0.05) and 48 hours (P≥0.1) incubation times. There were significant differences between isolates for the 4-hour (P≤0.001), 6-hour (P≤0.005) and 14 hour (P≤0.001) incubation times. Isolates in the 4, 6- and 14-hour incubation times were placed into 8, 7 and 9 groups, respectively (Table 2.3). The relative growth between strains over time is shown in figure 2.6. The combined data for the isolates placed the isolates into 9 groups when grown in 50% PEG and 5 groups when grown in 60% PEG (Table 2.3). This showed variability in PEG solutions used between isolates.

Table 2.3 Statistically significant groups of isolates grown in 50% PEG, 60% PEG or the combined data. The groups are listed in ascending order from the least growth (group 1) to the most growth (group 10). The red isolates are from the dry sites and the blue from wet sites.

	Growth in 50% PEG			Growth in 60% PEG			Combined 50%	Combined 60%
Group	6	14	48	4	6	14		
1	45,46,47, 48,49, 50, 51, 53, 54	31, 35, 38, 40, 41, 42, 43 44, 48, 50, 51, 53,54,	35, 36, 37, 39, 40, 41, 44, 45, 46, 47, 49, 50, 53	31, 32, 33,34, 35, 36 , 37	32, 33, 35 36, 37, 38, 41, 42, 44, 45, 46, 47, 48, 49, 50, 51, 52 53 , 54	32, 33, 45 52, 53, 54	35, 45, 47, 48, 49, 50, 51, 53	31, 32,33, 35, 36, 37, 38, 44, 45 47, 48, 50, 51,52, 53, 54
2	32	45, 47	42, 51, 52, 54,	49	31, 39	31	36, 46, 54	34, 40, 43, 46, 49
3	52	46, 49	48	38	40, 43	34	37, 40, 42, 44, 52	39
4	31, 37	36	33	52	34	35, 37, 39, 40, 49, 51	41	41
5	38	39, 52	32	39, 41, 42, 43, 44, 47, 51, 53, 54		50	43	42
6	35, 36, 42, 43, 44,	37	34	45		38	38	
7	40	32, 33, 34	43	46, 48, 50		36	31,33, 39	
8	33, 34		31	40		41, 42, 43, 46, 47, 48	32	
9	41		38			44	34	
10	39							

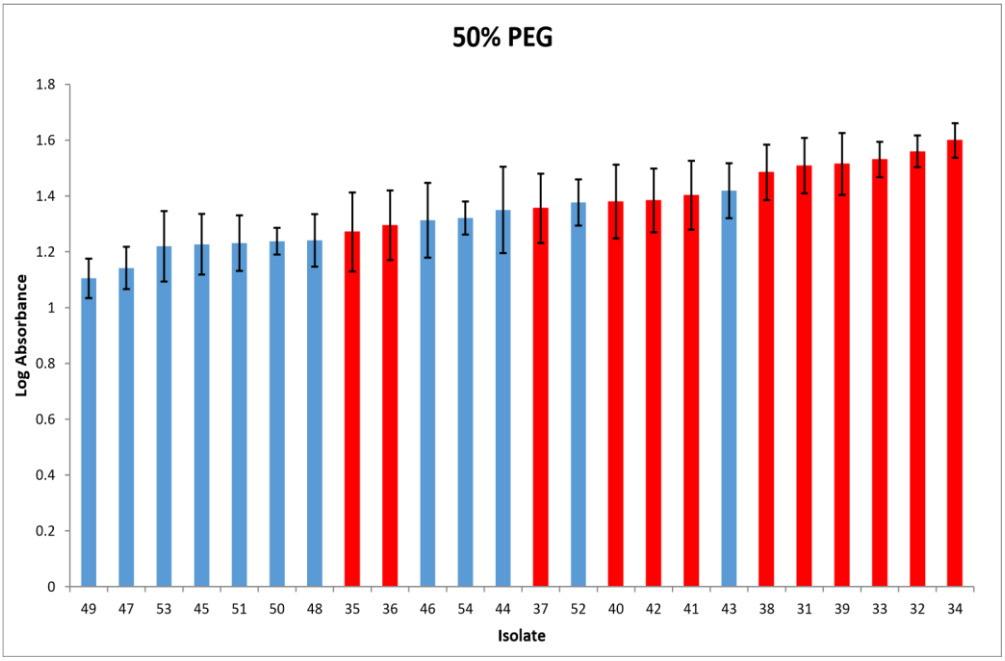


Figure 2.5 Relative absorbance for the isolates incubated in PEG with a 50% concentration. The graph is in a log scale. The red bars represent the two dry sites, site 6 (isolates 31-36) and site 17 (isolates 37-42). The blue bars represent the two wet sites, site 12 (isolates 43-48) and site 26 (isolates 49-54).

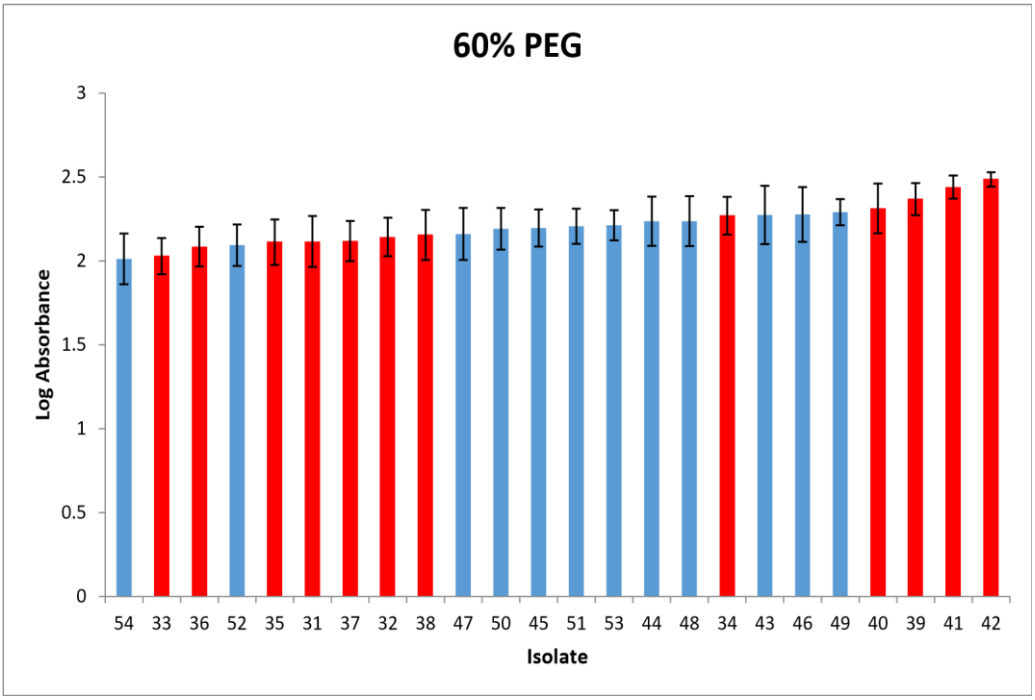


Figure 2.6 Relative absorbance for the isolates that were incubated in 60% PEG. The graph is in a log scale. The red bars represent the two dry sites, site 6 (isolates 3136) and site 17 (isolates 37-42). The blue bars represent the two wet sites, site 12 (isolates 43-48) and site 26 (isolates 49-54).

When the data was analysed collectively for the 50% PEG incubation there was a significant difference ($P<0.001$) between sites. The least growth was by isolates from site 26, followed by sites 12, 17 and 6 in ascending order.

When the data was analysed collectively for the 60% PEG incubation there was a significant difference ($P<0.001$) between sites. The sites were placed into three groups. The least growth was by isolates from sites 6 (dry site) and 26 (wet site), followed by site 12 (wet site), with the most growth from isolates from site 17 (dry site).

2.4 Discussion

The overarching goal of this chapter was to select genetically diverse isolates from sites that differed in the annual soil moisture deficit and test them for traits associated with desiccation tolerance. Other factors such as fertiliser and soil type were not considered in site selection as there was not enough time to include other variables, such as edaphic factors, into the experimental procedures. Two of these sites were from areas that were considered to have low soil annual moisture contents (100-140 days SMD). The other two sites were from areas that were considered to have high annual moisture contents (< 5 days SMD). The use of a random number generator to select 6 isolates from each of the sites removed any bias from the isolate selection process.

2.4.1 Genotyping

The ERIC PCR showed that there was genetic diversity in the isolates with 11 genotypes evident in the 24 isolates. Although some genotypes were site specific, others were found across all collection sites. This indicated that there was no obvious difference in genetic diversity between sites with high soil moisture, and low soil moisture deficits. There was an agreement with other studies showing a broad distribution of genotypes in New Zealand (Seehaver 2014, De Meyer, De Beuf et al. 2015). This has also been shown by Nangul, Moot et al. (2013) for the legumes *Trifolium repens* L , *Trifolium subterraneum* L. , *Trifolium glanduliferum* Boiss and throughout the world (Martínez-Romero and Caballero-Mellado 1996). The genetic diversity was done with ERIC-PCR which is commonly used to identify genetic variation in bacterial populations. Enterobacterial repetitive intergenic consensus polymerase chain reaction (ERIC- PCR) amplifies extragenic repetitive DNA sequences (Versalovic, Koeuth et al. 1991) and interspersed repetitive DNA (Versalovic, Schneider et al. 1994). Although these primers showed a good level of diversity it is possible that isolates with similar genetic fingerprints may not be clonal (Kosek, Yori et al. 2012). ERIC-PCR is easy to, but it produces a limited number of bands (Leung, Mackereth et al. 2004) which may make it difficult to differentiate between very similar strains. Close similarity of strains in NZ may be more likely than in other countries due to the widespread use of commercial inoculant strains. Other methods could have been used to create a genetic fingerprint, such as

amplified fragment length polymorphism. This method may have been a better choice since it would target more of the *Rhizobium*' genome, which would produce more bands giving a more accurate genotype (Leung, Mackereth et al. 2004). However, it is substantially more time consuming.

The phylogeny produced from the 16S sequences also indicated genetic some diversity amongst isolates. The main difference was isolate 52 was not a *Rhizobium sp*. but a massilia sp. with minor variation in the sequence obtained. Variation in 16S sequence of *R. leguminosarum* has also been shown by Menna, Hungria et al. (2006). However, there was no grouping between the ERIC-PCR fingerprint and the 16S sequences. This is expected since the 16s PCR produces a single band that targets conserved DNA sequences. The 16s is designed for species identification (Jensen, Webster et al. 1993). The ERIC-PCR showed a genotype within the species (Cubero and Graham 2002). There was also genetic diversity between the four sites although some genotypes were unique to a site such as isolate 42 has a unique genotype. It is well known that native strains of rhizobia are genetically diverse (Duodu, Brophy et al. 2009). Sampling a greater number of isolates would help improve understanding the genetic variation between and within each of these sites around New Zealand. Using a larger sample size from the collection these isolates were collected from, both number of isolates and sites collected from would give a better understanding of how the isolates can be characterised as desiccation tolerant.

PCR is an ecological tool to use to help determine the establishment and persistence of *Rhizobium* strains introduced into the New Zealand soils. Examples to PCR use are 16s rRNA (Jensen, Webster et al. 1993) which identifies species via targeting housekeeping genes, and enterobacterial repetitive intergenic consensus (ERIC) (Versalovic, Schneider et al. 1994). ERIC PCR allows short repeats to be identified. This allows the rhizobia to be categorised into phenotypes.

2.4.2 Bioassays

Two assays were developed to characterize traits associated with desiccation tolerance. These were the ability to form a biofilm and ability to grow in polyethylene glycol. The biofilm assay was chosen to give an indication of how much polysaccharide was produced in the various incubation times.

2.4.2.1 Biofilm production

Polysaccharide production has been shown to help prevent water loss (Potts 1994) by producing a biofilm surrounding the bacteria and numerous other important roles that help influence bacterial survival during desiccation stress (Engelhard 2004, Reina-Bueno, Argandoña et al. 2012).

The assay measured polysaccharide production at 24 and 48 h and these are similar to the time points used by Bomchil, Watnick et al. (2003) and Russo, Williams et al. (2006). Biofilms have been measured at 24 hours by Fujishige, Kapadia et al. (2006) and Russo, Williams et al. (2006). (Bomchil, Watnick et al. 2003) found that *Vibrio cholerae* produced maximum biofilm at 48 hours.

At both 24 and 48 h some isolates produced no measurable biofilm (not different to the uninoculated control). This may suggest that they are poor biofilm producers. Variation in the ability to form has been shown by Fujishige, Kapadia et al. (2006). However, the variations may have also been due to experimental error. The methodology required several rinsing steps and these steps may have either removed too much (disrupting the biofilm) or too few of the polysaccharides that were not bound on the well walls. Larger polysaccharide clumps can be removed during the washing process (Fujishige, Kapadia et al. 2006). It is possible that by extending the incubation time these isolates may have produced biofilm. Lack of measurable biofilm being due to the experimental procedure is supported by the results as group 1 contained different isolates at the 24 and 48 h incubation. This may have meant that the crystal violet stained residual polysaccharides and created erroneous absorbance measurements. This test was done in vitro which may influence biofilm formation as well as the biofilm could form differently depending on the complex medium it is grown in (Russo, Williams et al. 2006).

The method could be improved by using a more specific stain for trehalose as crystal violet is a non-specific or indirect stain for polysaccharides (Burton, Yakandawala et al. 2007). Also, the broth culture did not induce any stress on the isolates and they may not have responded as they would in soil. For a more accurate characterisation the isolates should also be grown in a stress inducing medium, such as polyethylene glycol amended broth, to determine if there are differences in extracellular polysaccharide production.

Although there was a significant difference between sites the isolates did not group by low or high SMD. This means that the ability of an isolate to produce extracellular polysaccharides does not appear to be determined by the site the isolate was recovered from. However, it is important to note that the i) isolates were not incubated in stressed conditions and ii) only a very small number of isolates were sampled. Zribi, Mhamdi et al. (2004) has shown that there is some variation with the genetic distribution of *rhizobia* at a species level. Some species will be found in all climates and other species can be found in niche climates i.e. tropical climates. There were statisical differences between the 24 hour and 48-hour incubation. A similar significant difference was observed by Bomchil, Watnick et al. (2003) for biofilm producted after 30 hours of incubation. This may be due to the growth relative growth rates between the isolates. For the combined data the incubation times were included together as a factor because of the differences between the statistical groupings for the individual isolates. By combining the data an it was more likely to

incorporate the relative growth rate of isolates and this was used to produce a final rank of strains for biofilm formation.

Some of the isolates produced large amounts of biofilms. Isolates such as 53 and 42 produced a consistently in the top half for high biofilm production. Since biofilm production has been attributed to having some tolerance towards desiccation (Chang, van de Mortel et al. 2007). Isolates that consistently produces high amount of biofilm can be selected as desiccation tolerance for further assays.

2.4.2.2 Polyethylene Glycol assay

The PEG assay was used to replicate a desiccation response. PEG is known to induce an osmotic stress response in bacteria (Cytryn, Sangurdekar et al. 2007) and the mechanisms involved with PEG survival are like those for desiccation stress tolerance (Radhouane 2007). Research has demonstrated that there is a lot of molecular cross-talk between the mechanisms of osmotic/salt and desiccation stresses (Mahajan and Tuteja 2005). In this study two concentrations of PEG were used, 50% and 60%. These have previously been used by McDougall (2015). In that work the 50% concentration applied a strong osmotic selection pressure and at 60% PEG was a concentration high enough to create a 99% kill rate. These two concentrations were therefore likely to simulate strong desiccation stress.

The results showed there was high variability in the absorbance data for each strain. To gain a more accurate result a greater number of replicates at each time point could be done. Variation may also have been introduced by “clumping” of the bacteria whch interfered with the absorbance measurements. Rhizobia are known to clump during growth and this may be exacerbated by polysaccharide production, which may have been highly induced under PEG stress. There were significant differences between the sites, but they were not placed into the same groups as for the biofilm formation. Overall, the results from the 50% PEG showed that a greater number of isolates from the dry sites were able to withstand and grow in higher PEG concentrations at the 50% PEG with 10 out of 12 dry site isolates were found in the top half of the survivability. Alternatively, the 60% PEG showed there was no grouping of strains. Approximately half the high ends were wet and dry site isolates (seven out of 12 were wet site isolates)

When the data from table 2 and 3 are compared a slight pattern emerged for several of the isolates. Some of the strains and genotypes align between the PEG and bioassay. An example of this are the isolates from genotype E, F, I, and J from figure 6. From the data there are 3 desiccation tolerant (isolates 42, 41, and 34) and 4 non tolerant (isolates 47, 50, 52 and 48) isolates. All four non-tolerant strains are from genotype E. The tolerant strains are from genotypes J, F, and I. Isolate 53 was carried through carried through as a tolerant strain as it was in the top half of 60% PEG and it was very high in all biofilm assays.

1.5 Conclusion

The strains of *R. leguminosarum* were genotypically diverse with several unique genotypes present. There was not enough evidence to conclude that, in general, isolates from a high soil moisture deficit are more desiccation tolerant than isolates from a low soil moisture deficit. More isolates would be required to determine whether that was the case. The isolates from the high soil moisture deficit did not produce more biofilm or survive longer in PEG. However, using a biofilm assay in conjunction with a PEG assay candidate isolates can be selected for desiccation tolerance. An additional assay may need to be added since some isolates showed tolerance (53) in chapter 4 but was not in in high concentration in the PEG assays.

Chapter 3 - Characterisation of genes involved in trehalose production by *Rhizobium leguminosarum*

3.1 Introduction

There are several mechanisms suggested for desiccation tolerance such as increased production of the polysaccharide trehalose. Trehalose has a significant effect on desiccation tolerance in other species, such as Cyanobacterium sp. and is therefore likely to be important in *R. leguminosarum*. The sequence of genes involved in the production of trehalose may vary between strains and are good targets genes for analysis as potential mechanisms of desiccation variation. The genes responsible for trehalose production have been characterised as participating in three biosynthetic pathways. The most common pathway used by bacteria contains genes that produce trehalose-6-phosphate synthase (*otsA*) and trehalose-6-phosphate phosphatase enzymes (*otsB*)(Engelhard 2004). The other two biosynthesis pathways for trehalose are also found in bacteria (Cytryn, Sangurdekar et al. 2007). The MOTS pathway consists of genes that encode the enzymes maltooligosyltrehalose synthase (*tre*Y) and maltooligosyltrehalose trehalohydrolase (*tre*Z). The third trehalose biosynthesis pathway is the *treS* pathway that involves a gene that encodes the trehalose synthase enzyme.

It is possible that rhizobia that have adapted to dry conditions have differences in their gene sequences. The differences within the genes may influence the efficiency in which the gene products function. Primers can be designed to target these genes to help determine if there are polymorphic differences between these genes in different strains. This aim of this chapter is to i) design primers to amplify genes involved in trehalose production by *R. leguminosarum*, ii) to investigate the location of the genes within the *R. leguminosarum* genome and iii) to sequence the genes involved in trehalose production from strains adapted to wet or dry soils. The chapter tests the hypothesis that "*Candidate genes associated with the ability of a strain to produce trehalose are polymorphic between wet adapted and dry adapted strains".*

3.2 Materials and methods

3.2.1 Primer design

3.2.1.1 Primer development

The candidate genes chosen are described in Table 3.1. All the genes are involved in trehalose biosynthesis.

Table 3.1 Genes involved in trehalose synthesis

Gene	protein	Function	References
otsA	Alpha, alpha-Trehalose6-phophate synthase [UDP forming...]	Conversion of UDPglucose and glucose into UDP and α,α trehalose-6-phosphate	(De Smet, Weston et al. 2000, Engelhard 2004, McIntyre, Hore et al. 2007, Suárez, Wong et al. 2008, Fernandez, Béthencourt et al. 2010, Reina-Bueno, Argandoña et al. 2012)
*ots*B	Trehalose-6-phophate phosphatase	Conversion of UDP and α,α trehalose-6phosphate into UDP, phosphate group and α,α-trehalose	(De Smet, Weston et al. 2000, Engelhard 2004, McIntyre, Hore et al. 2007, Suárez, Wong et al. 2008, Fernandez, Béthencourt et al. 2010, Reina-Bueno, Argandoña et al. 2012)
*tre*Y	Maltooligosyl trehalose synthase	Terminal end (4 end Carbons with a α 1-4 linkage) of Sugar chain into α 1-1 linkage maltooligosyltrehalose	(De Smet, Weston et al. 2000, Engelhard 2004, McIntyre, Hore et al. 2007, Suárez, Wong et al. 2008, Fernandez, Béthencourt et al. 2010, Reina-Bueno, Argandoña et al. 2012)
*tre*Z	Maltooligosyl trehalose trehalohydrolase	Cleavage of terminal disaccharide, releasing trehalose	(Maruta, Hattori et al. 1996, De Smet, Weston et al. 2000, Simon-Rosin, Wood et al. 2003, Engelhard 2004, Suárez, Wong et al. 2008, Fernandez, Béthencourt et al. 2010, Reina-Bueno, Argandoña et al. 2012)
treS	Trehalose synthase	Transglucosylation of maltose into trehalose.	(De Smet, Weston et al. 2000, Engelhard 2004, McIntyre, Hore et al. 2007, Suárez, Wong et al. 2008, Fernandez, Béthencourt et al. 2010, Reina-Bueno, Argandoña et al. 2012)

Rhizobium leguminosarum genomes were downloaded from the JGI database (http://jgi.doe.gov/). The genomes downloaded were of the commercial strains of *R. leguminosarum*, WSM1689, cc278f and TA1.

A copy of each of the genes was obtained from GenBank. The gene sequences were obtained from *R. leguminosarum* bv trifolii CC278f (2509276052), *R. leguminosarum* bv trifolii WSM1689 (2510065019), *Rhizobium leguminosarum* bv *trifolii* TA1 (2510461076) and *Rhizobium leguminosarum* bv *viciae* UPM1137 (2513237085) genomes from the Joint Genome Institute (https://jgi.doe.gov). These genes were used to find the gene homologue in *R. leguminosarum* using a BLASTn search. Once located, the neighbouring genes in *R. leguminosarum* (three genes on either side of the target gene, there relative size and orientation) were also identified. Each gene from the *R. leguminosarum* strains was aligned using ClustalW (MEGA 5.0). Where the DNA sequences were identical 20 bp forward and reverse primers were developed to be as close to the edge of the gene as possible. Four forward and four reverse primers were developed for each gene to produce a product < 2500 base pairs. If the gene was longer than 2500 base pairs overlapping primers were developed. Where the primer sequence was not identical, a degenerate base notation was used according to standard nomenclature IUPAC nucleotide code[2] (appendix 7.3.1). The primers were required to have < 3 degenerate bases. The melting temperature of the primers was approximately 60°C. The primers were also tested for selfcomplementarity and complementarity to each other. Any designed with high complementarity were discarded. Primers were ordered from IDT and diluted to a working solution of 10 ng/µL.

3.2.1.2 DNA extraction

A bacterial broth was grown by placing 1 mL of yeast mannitol broth (YMB) in a sterile 1.7 mL tube and inoculating with a single bacterial colony from isolates 32, 40, 41, 42, 43, 45, 47, 50, 51, and 53 (using commercial strains WSM1325, cc275e and TA1 for optimisation). This was placed in a shaking incubator at 28°C for 24 h. Control tubes were included to check for contamination. DNA was extracted from broth cultures using the PUREGENE DNA extraction kit (Qiagen) according to the manufacturer's instructions. The purified DNA was resuspended in 30 µL of sterile water and stored at 4°C. The DNA concentration was quantified in a Nanodrop spectrophotometer and adjusted to 10 ng/µL for PCR.

3.2.1.3 PCR

PCR was done to amplify the genes for sequencing so that polymorphism between isolates could be detected. A master mix was made from each ingredient shown in Table 3.2. An aliquot of 19 µL of the master mix was added to each of the labelled tubes. An additional 1 µL of sample DNA was added to the appropriate tubes and pipetted to mix (including controls). The tubes were placed in the Veriti thermal cycler and amplified with optimised conditions, conditions differed between genes.

[2] https://www.bioinformatics.org/sms/iupac.html

Table 3.2 PCR Master Mix

PCR MASTER MIX COMPONENTS**REAGENTS**	REACTION MIX × 1 (µL)
Dream Taq master mix	10
Degenerate forward primer (10 µM)	1
Degenerate reverse primer (10 µM)	1
H_20	7
Total	19
Sample DNA	1 (10 ng)

3.2.1.3 PCR optimisation

The amplification reaction for the primers was optimised for annealing temperature. To do this an annealing gradient below 60°C of four different temperatures within 6°C between the highest and lowest annealing temperatures were assessed. DNA extracted from the commercial strains, WSM1325, cc275e, and TA1 were used for PCR optimisation. The extension time was altered in accordance with the expected product size. This was calculated on the basis of 1 minute per 1000 bp. The thermal cycler was set for 94°C for 3 min, followed by 35 cycles of 94°C for 30 s, 54°C for 30 s (other annealing temperatures were 56 °C, 58 °C and 60 °C) and 72°C for 1 min (adjusted for products that were expected to be larger) and a final extension of 72°C for 7 min.

3.2.1.5 Gel electrophoresis

The PCR product was run on a 1% agarose gel (1 g agarose: 100 mL water). Six □L of each sample (including controls) were added to each well. Each of the samples was loaded into wells on the gel submersed in 1x TAE buffer. 1 KB plus DNA ladder (Invitrogen) was mixed with a 6x loading dye (2µL ladder with 3 µL loading dye) and placed into one well (5 µL) per gel for size comparison. The samples were separated at 120V/cm for 45 min. The gels were stained with ethidium bromide for 10 min, rinsed in sterile water and photographed under UV light.

3.2.1.6 Sequencing

The PCR products were sequenced directly by the Bio-Protection Research Centre Sequencing Facility, Lincoln University, NZ in both direction using primers specific for the target gene. The resultant sequences were edited using DNAMAN Version and Chromas lite to remove any poor or ambiguous sequence. The DNA sequences were compared to other known sequences in GenBank using the BLASTn algorithm on NCBI (http://blast.ncbi.nlm.nih.gov/) to confirm the sequences were the target genes. Gene polymorphisms were identified by alignment of the target genes from all strains.

3.3 Results

3.3.1 Gene contigs

Analysis of the genomes from strains WSM1689, cc278f and TA1 indicated that the neighbouring genes were variable between strains (Figure 3.1 and 3.2).

3.3.2 Primers

Primers were designed to each of the genes as follows:

- *ots*B - two forward primers were developed (TP7F- 5' ACGACGACGCTACTCAGGA 3' & TP41F-
  5' TCTTGGAAGAGCCCGACCA 3') and two reverse primers (TP743R- 5' GCAATSACGTTTCTRACCAG 3' & TP653R- 5' CCACGCGCATTGGCGATC 3').
- *Tre*S - five primers were designed, two that flanked the gene TS130F- 5' CAGAAGCTCGATCATATCGC 3' \and TS1642R- 5'GGAACCAGAAGAAACCGTAG 3 ', and two internal to those namely TS155F- 5' TCGGCGTCAATGCCATCTG 3' and TS1596R- 5'TCGGCCGATCGGCGGAAA 3'.
- *ots*A - two forward primers (ATPS10F- 5' CTTRTYGTCGTTTCMAATCG 3' & ATPS86F- 5' TGCAGGCMGCYCTSCAGG 3') and two reverse primers (ATPS1344R- 5' GCACCARMGWGAGACGTC 3' & ATPS1246R- 5' CATTGGCRGTGCCTTCGAC 3') were developed.
- *tre*Y - Two forward (MOTS11F-5' CGACAGCGACCTACCGGA 3' & MOTS142F-5' TCGACCCACGGCTATGACG 3') and two reverse (MOTS2091R-5' CTTGAGCAGYGTCTGCGAC 3' &MOTS2585R-5' GGCTGGCTCCCGAGCAG 3') primers were developed.
- *tre*Z - two forward (MOTT117F- 5' TCCCYTGCATGAAAGCGTG 3' & MOTT312F- 5' CGAAGTGGTSGAYCTTTCC 3') and two reverse (MOTT1633R- 5'CCGAGCGATAGACCGCGT 3' & MOTT1756R- 5' GGAAGATCGTCTCGCCRTC 3') primers were developed.

For each primer combination the approximate amplicon size was determined as per table 3.1.

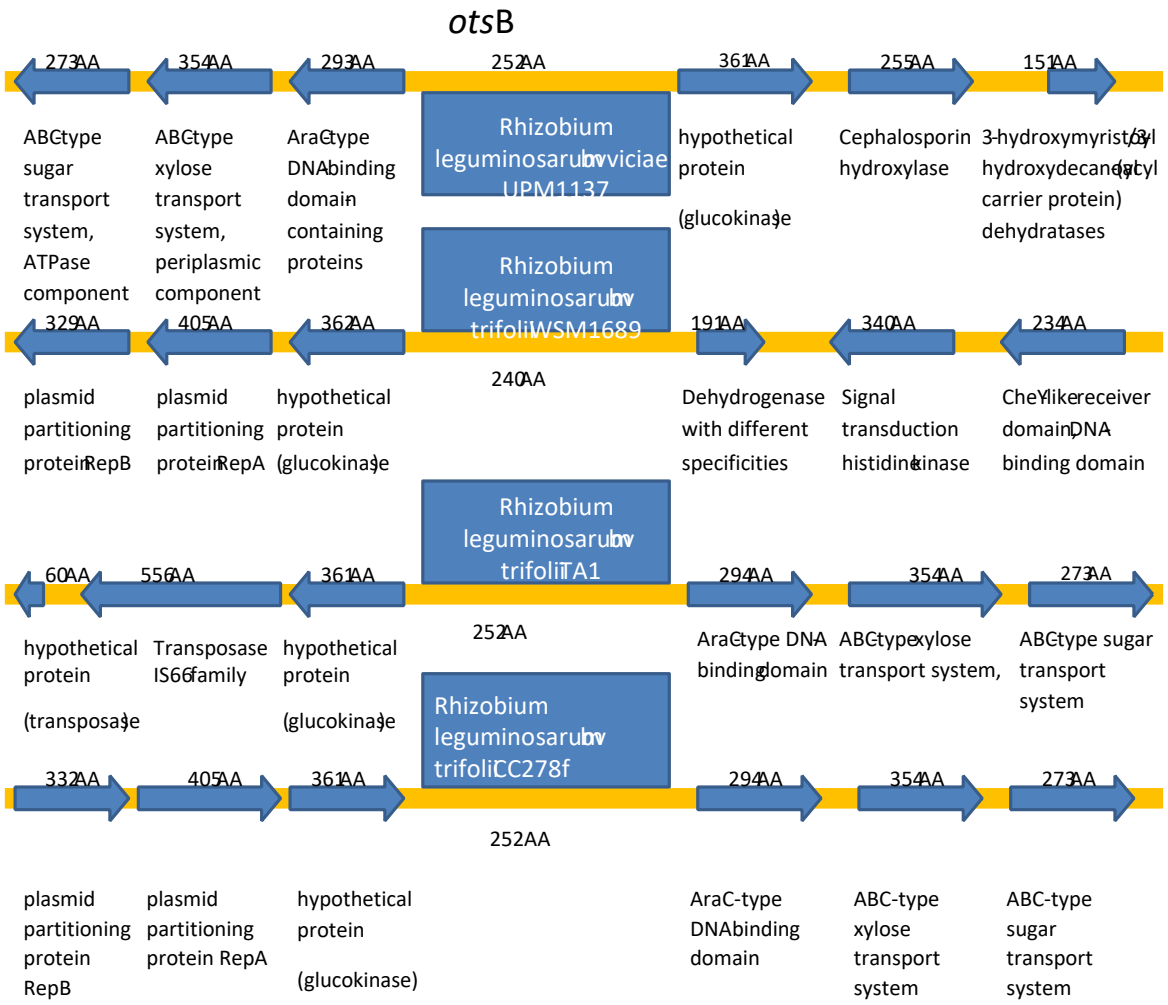


Figure 3.1 This figure shows the trehalose-6-phosphate phosphatase (otsB) gene contiguous sequence in four strains of *R. leguminosarum*. The orientations of the genes are illustrated by the arrow direction, right facing arrows are on the 5'- 3' sense strand and left pointing arrows are located on the antisense strand. The relative peptide strand is also illustrated.

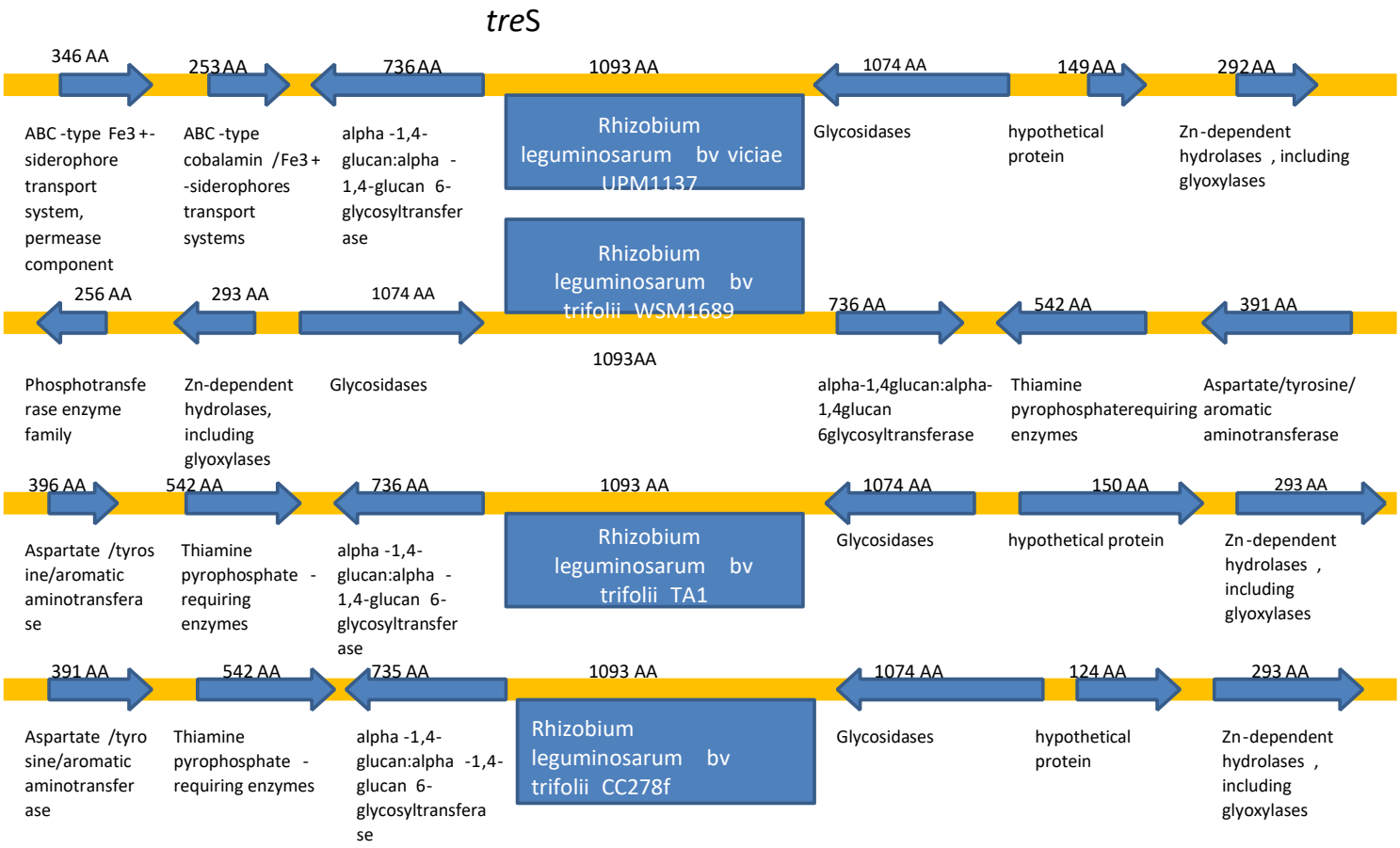


Figure 3.2 This figure shows the trehalose synthase (treS) gene contiguous sequences for four strains of R. *leguminosarum*. The orientations of the genes are illustrated by the arrow direction, right facing arrows are on the 5'- 3' sense strand and left pointing arrows are located on the antisense strand. The relative peptide strand is also illustrated.

Table 3.3 This table listed the different primer combinations for each of the target genes.

Set	Primer Combination	Target Gene	Gene Pathway	Expected Amplicon size (bp)
A	TP7F & TP653R	*ots*B	TPS-TPP	646
B	TP7F & TP743R	*ots*B	TPS-TPP	736
C	TP41F & TP653R	*ots*B	TPS-TPP	613
D	TP41F & TP743R	*ots*B	TPS-TPP	703
E	ATPS10F& ATPS1344R	*otsA*	TPS-TPP	1334
F	ATPS10F & ATPS1246R	*otsA*	TPS-TPP	1246
G	ATPS86F & ATPS1344R	*otsA*	TPS-TPP	1258
H	ATPS86F & ATPS1246R	*otsA*	TPS-TPP	1160
I	MOTS11F & MOTS2091R	*tre*Y	TreY-TreZ	2080
J	MOTS11F & MOTS2585R	*tre*Y	TreY-TreZ	2574
K	MOTS142F &MOTS2091R	*tre*Y	TreY-TreZ	1949
L	MOTS142F & MOTS2585R	*tre*Y	TreY-TreZ	2443
M	MOTT117F & MOTT1633R	*tre*Z	TreY-TreZ	1518
N	MOTT117F & MOTT1756R	*tre*Z	TreY-TreZ	1640
O	MOTT312F & MOTT1633R	*tre*Z	TreY-TreZ	1322
P	MOTT312F & MOTT1756R	*tre*Z	TreY-TreZ	1444
Q	TS130F & TS1642R	*treS*	TreS	1512
R	TS130F & TS1596R	*treS*	TreS	1466
S	TS155F & TS1642R	*treS*	TreS	1482
T	TS155F& TS1596R	*treS*	TreS	1441

3.3.2 Primer optimisation

Each of the primer combinations in table 3.1 were tested and optimized using three commercial strains, WSM1325, cc275e and TA1. At least one of the commercial strains produced bands of the expected size for each primer combination in all primer combinations.

Trehalose-6-phophate phosphatase (*otsB*) - All primer combinations produced bands for all three commercial strains (figure 12) of the expected size. However, bands in addition to the expected amplicon were produced in primer combinations A, C and D for some of the strains for strain WSM1325.

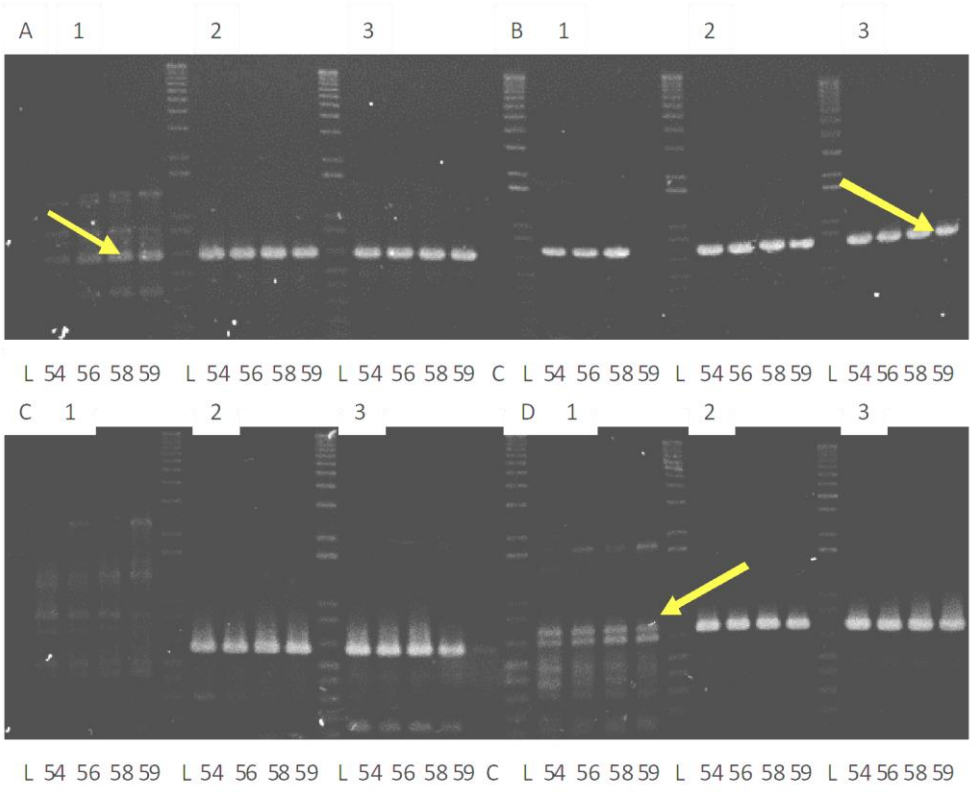


Figure 3.3 1% agarose gel amplicons for targeted gene otsB with primer combinations A, B, C, and D. L = 1kB DNA ladder. 1, 2, 3 are commercial strains WSM1325, cc275e and TA1, respectively. The numbers at the base of each gel are the annealing temperatures. The yellow arrows indicate the expected products.

α, α-Trehalose-6-phophate synthase (*otsA*) – Primer combination E produced two very faint bands at 56°C annealing temperature for all three commercial strains. One of these two bands was the target size of ~1334 bp. Primer combination F produced bright bands at the target size in all four annealing temperatures, cc275e produced a second faint band between 2,000- 3,000 bp at some annealing temperatures. Primer combination G produced amplicons of the target size at most annealing temperatures for strains cc275e and TA1. Strain WSM1325 produced distinct bands in all four annealing temperatures but these were ~650 bp, which was not the target size. Combination H produced amplicons of the

expected size for all three commercial strains in all annealing temperatures used with WSM1325 producing multiple bands.

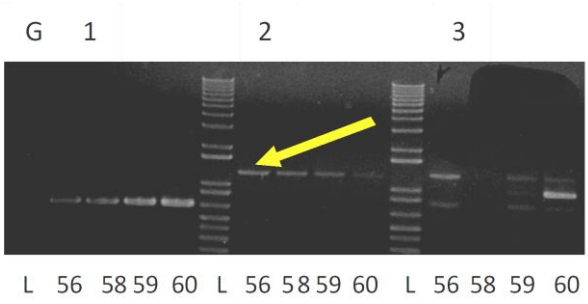


Figure 3.4 1% agarose gel of amplicons of target gene otsA produced by primer combination G. L = 1kB DNA ladder. 1, 2, 3 are commercial strains WSM1325, cc275e and TA1, respectively. The numbers along the bottom are the annealing temperatures. The yellow arrows indicate the expected products.

Trehalose synthase (*treS*) – A single amplicon was produced in all four primer combinations (Q, R, S and T; Figure 14). The annealing temperature of 60°C for WSM1325 did not produce a band in primer combination T.

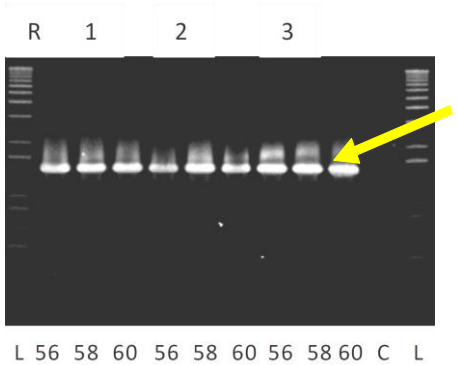


Figure 3.5 1% agarose gel of amplicons for target gene treY produced by primer combination R. L = 1kB DNA ladder. 1, 2, 3 are commercial strains WSM1325, cc275e and TA1, respectively. The numbers along the bottom are the annealing temperatures. C = negative control.

Maltooligosyl trehalose synthase (*treY*) – Bands of the expected size were produced in all four primer combinations (I, J, K, and L) in all three commercial strains and at all annealing temperatures. However, all primers produced additional bands for most strains and annealing temperatures combinations (Figure 3.6).

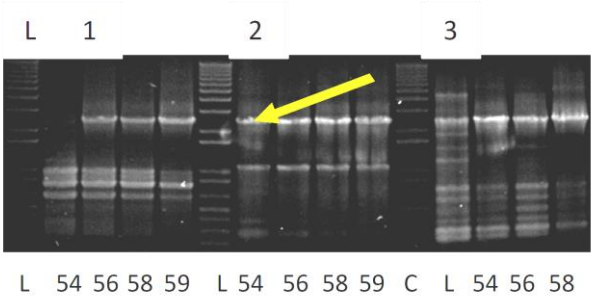


Figure 3.6: 1% agarose gel of amplicons for target gene *treZ* produced by primer combination L. L = 1kB DNA ladder. 1, 2, 3 are commercial strains WSM1325, cc275e and TA1, respectively. The numbers along the bottom are the annealing temperatures. The yellow arrows indicate the expected products.

Maltooligosyl trehalose trehalohydrolase synthase (*treZ*) - All primers combinations (M,N,O,P) produced bands of the expected size (Figure 16). Some of the annealing temperatures produced additional bands for strains WSM1325 and TA1.

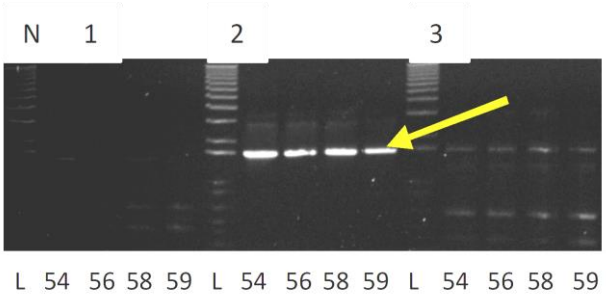


Figure 3.7 1% agarose gel of amplicons for target gene treS produced by primer combination N. L = 1kB DNA ladder. 1, 2, 3 are commercial strains WSM1325, cc275e and TA1, respectively. The numbers along the bottom are the annealing temperatures. The yellow arrows indicate the expected product.

3.3.3 Trehalose biosynthesis gene amplification

The best primer and annealing temperature were selected for each of four genes (Figure 17) and these were:

- For trehalose-6-phosphate phosphatase (*otsB*) primer combination B at an annealing temperature of 59^{o}C produced a target band for all 10 isolates of approximately 700 bp.
- For α, αtrehalose-6-phophate synthase (*otsA*) primer combination F at annealing temperature 59^{o}C produced bands of the target size of 1250 bp for all the isolates used but isolate 47 which produced an additional band that was <400 bp.
- For maltooligosyl trehalose synthase (*treY*) isolates 42, 47, 53, cc275e and TA1 produced bands at the target size of ~2450bp for primer combination L at an annealing temperature a 59^{o}C. Isolates 53 and cc275e produced an additional band of 1000-1600 bp in size. Isolate TA1 produced an additional band of <400 bp. Isolate 41 did not produce any bands.
- For maltooligosyl trehalose trehalohydrolase (*treZ*) primer combination N all annealing temperature 59^{o}C produced bands of the expected size for isolates 42, 53, cc275e and TA1. TA1 produced an additional two bands <800 bp. Isolates 41 and 47 did not produce any bands.
- For trehalose synthase (*treS*) primer combination R at annealing temperature 59^{o}C produced bands at the target size for isolates 40, 41, 42, 43, 47, 50, 51, and 53.

Isolates 40, 41, 42, and 45 produced an additional band smaller than 300bp. Isolates 32 and 45 did not produce bands at the target size.

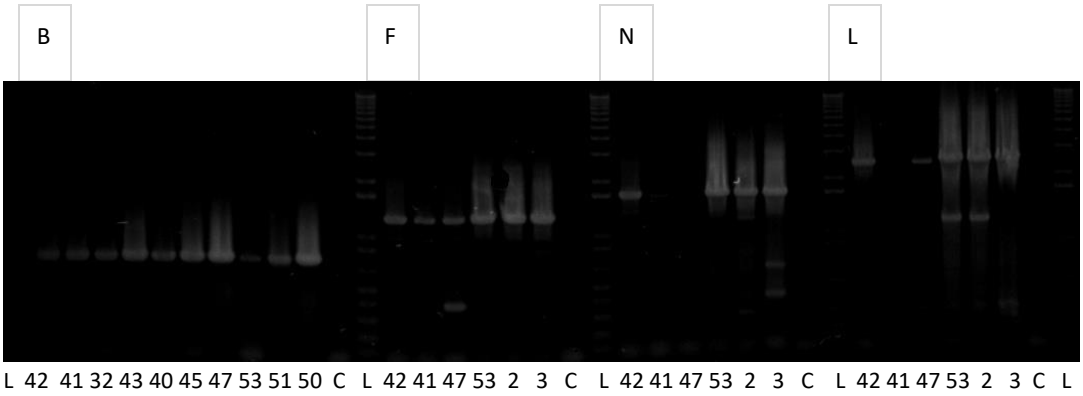


Figure 3.8 1% agarose gel of target gene amplicons primer combinations F, L, N and R (below the gel) = 1kB DNA ladder. 2 and 3 are commercial strains cc275e and TA1. The other numbers along the bottom are the isolates.

The band of the expected size was excised and purified for primer combinations F, L, N and R for four isolates 41, 42, 47, and 53 (all 10 isolates were used for primer combination R) and commercial strain cc275e and TA1. Purified amplimers are shown in Figure 3.8.

No amplimers were successfully purified from primer combination F. Multiple bands for all isolates that were <650 bp were the results of purification of combination N with isolate 53 was the only isolate to retain a band at the target amplicon size (1640bp). For amplimers from primer combination L only isolates 53, 47, and 42 retained bands at the target amplicon size of 2443 bp. Bands <300 bp was also co purified.

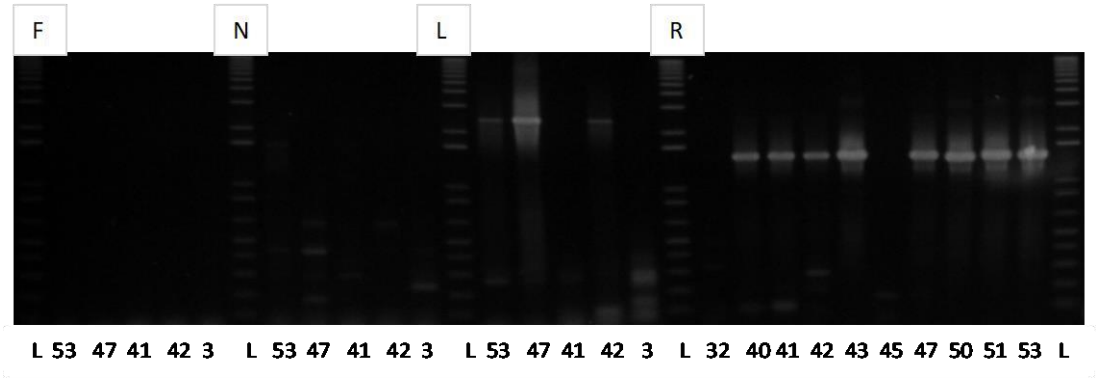


Figure 3.9 1% agarose gel of target gene amplicons retained after gel purification. The letters F, N, L and R are the primer combination. L (below the gel) = 1kB DNA ladder. 3 is commercial strain TA1. The numbers along the bottom are the isolates.

3.3.4 Sequencing

Nine of the 10 *otsB* sequences amplified and sequenced produced high quality electropherograms. Some of the commercial strains for each primer combinations were also sequenced. These sequences were placed into six groups, three of which are subgroups (figure 3.3). There was a 95% identity between the 9 isolates, 32, 40, 41, 42, 43, 47, 50, 51, and 53. There were a total of 74 single nucleotide polymorphisms between the genes form the nine isolates. A translation of the nucleotide sequence showed a 98% identity between sequences with 8 amino acid polymorphisms. Of these six were conservative substitutions that did not change the properties of the amino acids. Another two of the amino acids substitution was a change in polarity and a change between a negative charge and polar amino acids figure table 8.

Isolates 40 and 51 were in a subgroup of isolate 32. Isolates 32, 40 and 51 were closest to the out group *treS*. Isolate 42 was placed in its own group. Isolates 43 and 47 were in a subgroup of isolate 50. Isolates 41 and 53 were placed in their own group. Comparison of the translated sequence with those present on GenBank identified the genes as trehalose6-phosphate phosphatase (*otsB*) that was either chromosomal or plasmid located (Table 3.4).

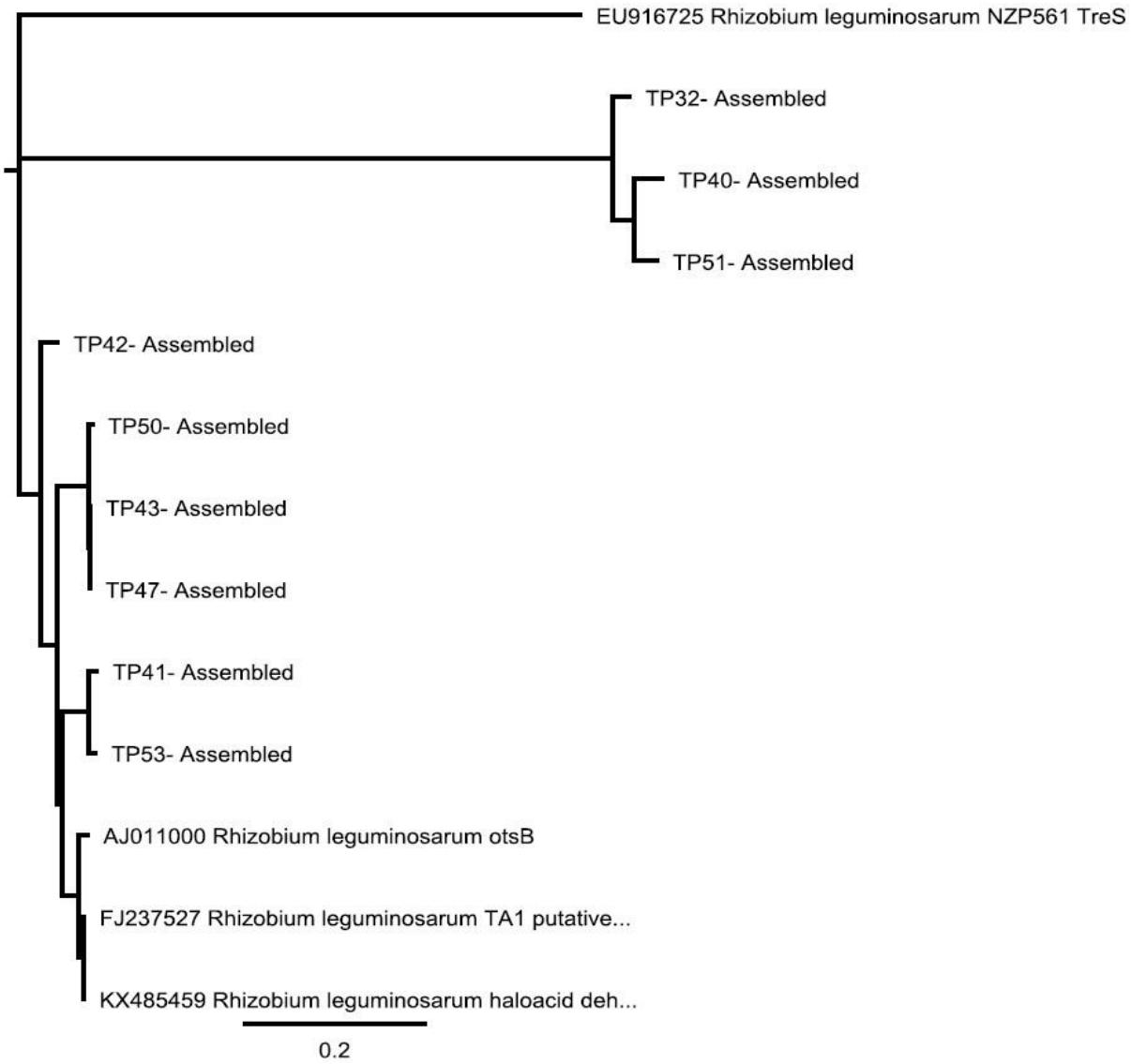


Figure 3.10 This is a rooted phylogenetic tree of nucleotide sequences. EU916725 *Rhizobium leguminosarum* NZP561 treS, FJ237527 *Rhizobium leguminosarum* TA1 putative substrate-binding component of ABC transporter gene, and KX485459 *Rhizobium leguminosarum* haloacid dehydrogenase are out groups. AJ011000 Rhizobium leguminosarum otsB is an annotated gene for *Rhizobium leguminosarum*. TP32- Assembled, TP40-Assembled, TP51-Assembled, TP42-Assembled, TP50Assembled, TP43-Assembled, TP47-Assembled, TP41-Assembled, and TP53Assembled are the isolates being studied.

Table 3.4 The translated BLASTx results of the sequence queries otsB.

ISOLATE	QUERY COVERAGE %	IDENTITY %	GENE	ACCESSION NUMBER
WSM1325	99	100	trehalose-phosphatase	WP_012760603.1
TA1	99	100	trehalose-phosphatase	WP_029767531.1
CC275E	99	100	trehalose-phosphatase	WP_033184280.1
40	99	100	trehalose-phosphatase	WP_033184280.1
32	99	100	trehalose-phosphatase	WP_029767531.1
51	99	99	putative trehalose-phosphatase (plasmid)	CAK10692.1
53	99	100	trehalose-phosphatase	WP_033184280.1
41	99	100	trehalose-phosphatase	WP_033184280.1
47	99	99	putative trehalose-phosphatase (plasmid)	CAK10692.1
43	99	99	putative trehalose-phosphatase (plasmid)	CAK10692.1
50	99	99	putative trehalose-phosphatase (plasmid)	CAK10692.1
42	99	100	trehalose-phosphatase	WP_029767531.1

The primer combinations targeting genes otsA, treY, treZ and treS did not produce clean sequences when amplimers were directly sequenced. Resequencing of the gel purification amplimers failed to improve sequences for *otsA, treY,* and *treZ*. However, the gel purification did produce better sequencing for the *treS* gene of four isolates 47, 51, 43, and 53. All isolates were identified as *tres*S by BLASTx sequencing with isolates 47 and 51 grouping together and isolates 43 and 53 grouping together with the EU916725 *Rhizobium leguminosarum* NZP561 *treS* (Figure 3.11). There was a total of 75% identity for the gene alignment. Isolate 47 was missing approximately 470 base pairs (from the 3' end 190 and 5' 280 nucleotides) from the aligned sequence. Isolate 41 was also missing approximately 70 base pairs from the 5' end and 100 base pairs at the 3' end. When the sequence was trimmed the identity was 97%.

Table 3.5 Amino acid substitutions in the otsB gene. • represent isolates that contain the substitution.

AMINO ACID NUMBER	AMINO ACID SUBSTITUTION	ISOLATES								
		32	40	41	42	43	47	50	51	53
64	Serine → Alanine	•			•					
73	Glutamic acid→ Alanine	•			•					
94	Valine → Isoleucine					•	•	•	•	
98	Alanine → Phenyalanine	•			•					
142	Aspartic acid→ Glutamic acid					•	•	•	•	
145	Histidine→ Arginine		•	•						•
150	Leucine → Valine → Isoleucine		•	•		••	••	••	••	•
213	Valine → Leucine					•	•	•	•	

Figure 3.11 A rooted phylogeny tree of the four treS nucleotide sequences, TS47Assembled, TS51-Assembled, TS43-Assembled, and TS53-Assembled. EU916725 *Rhizobium leguminosarum* NZP561 treS is the characterised gene found in *Rhizobia leguminosarum*. AJ011000 *Rhizobium leguminosarum* otsB is an out group from another trehalose biosynthetic pathway.

3.4 Discussion

This chapter attempted to amplify five genes involved in the biosynthesis of trehalose from the strains of *R. leguminosarum* isolated from soils with differences in their SMD. It was hypothesised that they may contain sequence polymorphism and that this may be a mechanism by which different amounts of the polysaccharide trehalose was produced, contributing to variation in desiccation tolerance soils. Analysis of the genomes from four strains present on the JGI database showed that the placement of the genes in the chromosome varied with difference between the contiguous sequences. Primers were developed that successful amplified the otsB and treS genes. Sequence polymorphism was observed for these genes between isolates.

Analysis of the four *Rhizobium leguminosarum* bv trifolii genomes on the JGI database for the placement of the five genes (trehalose-6-phosphate phosphatase, α, α trehalose-phosphate synthase, maltooligosyl trehalose synthase, maltooligosyl trehalose phosphatase and trehalose synthase) encoding enzymes associated with the biosynthesis of trehalose showed that the location of the genes on the chromosome varied between *R. leguminosarum* bv *trifolii* cc278f, *R. leguminosarum* bv trifolii WSM1689, *R. leguminosarum* bv *trifolii* TA1and *R. leguminosarum* bv *viciae* UPM1137 strains. This is illustrated by figure 3.1, 3.2 and appendixes 5.1, 5.2, and 5.3. Of the five trehalose biosynthetic pathways, three have been identified *ots*A-*ots*B, *tre*Y-*tre*Z and treS pathways in rhizobia. All five of these genes belonging to the three pathways have been identified in *Rhizobium* sp. (Engelhard 2004, Cytryn, Sangurdekar et al. 2007, Reina-Bueno, Argandoña et al. 2012). These three pathways have been described in numerous bacteria to date, more specifically the Rhizobiaceae family (McIntyre, Davies et al. 2007). Since the trehalose biosynthetic pathways are widely distributed in nature (Avonce, Mendoza-Vargas et al. 2006) and it has numerous roles associated with stress response (Potts 1994) it was a good set of candidate genes to explore.

Alignment of the gene homologues from *R. leguminosarum*, WSM1689, TA1 and cc278f allowed the design of 20 bp primers. Twenty nucleotides should have been enough to ensure specificity for annealing to the target genes (Lang and Orgogozo 2011). The primers were also designed to have a GC clamp at the 3' end of the primer to help promote tight binding and improve the efficiency of

extension by the polymerase (Dieffenbach, Lowe et al. 1993). Runs of three or more G's or C's were avoided to prevent mispriming (Innis and Gelfand 1999). An allowance of up to three degenerate bases was acceptable (Kwok, Chang et al. 1994). The addition of the degenerate bases will cover all possible variants of the target gene sequence (Jabado, Palacios et al. 2006, Lang and Orgogozo 2011). To make sure the primers were specific for the target genes the primers were used to search the GenBank database. Despite these aspects of primer design nonspecific amplification was still observed after PCR for many of the primer combinations. This may have been because several biovars were used. Another reason why multiple bands may have been produced is because some of the genomes had multiple copies of the genes as the contigs figure 10, 11 and appendixes 22, 23, and 24 illustrates. There may have been repeats within the gene sequences that created smaller bands (Kalendar, Lee et al. 2009). There was degeneracy in some of the primers however primers that had no degeneracy also produced multiple bands. Some of the amplicons had a large target size and this may have contributed to multiple amplicons being formed. The size of the target amplicon (those >2000bp) may have influenced the efficiency of the Taq polymerase causing it to make multiple products. Although Taq Polymerase can amplify 2kb/min (Innis and Gelfand 1999) it may have may not have had long enough to extend accurately or it may have required a higher extension temperature to reduce mismatching (Ford and Rose 1995).

When designing the primers, it was observed that the DNA sequence of the genes was variable between biovars. There were observable variations between biovars when developing the primers. When initially developing the primers a fifth strain was selected
(*Rhizobium phaseoli*). The variation between the five strains was too high to create primers (data not included). The strains did not have an observable difference between the viciae biovar and trifolii biovar. However, McIntyre, Davies et al. (2007) found that there is a difference in location within the bacterial cells, which was shown by table 3.3.4.1. The blast search for the sequences identified that both plasmid and chromosomal locations for the genes were possible.

During the primer optimisation processes a single variable was manipulated, annealing temperature. The annealing temperature was the only factor that was manipulated because it has been shown to increase specificity of amplification

(Sipos, Székely et al. 2007). The annealing temperature was also used because it is simple to manipulate. The size of the target amplicon (those >2000bp) may have influenced the efficiency of the Taq polymerase causing it to make multiple products. Although Taq Polymerase can amplify 2kb/min (Innis and Gelfand 1999) it may have may not have had long enough to extend accurately. This may have also increased the error rate (Eckert and Kunkel 1991). Manipulating the temperature of the extension time (mainly for treY) may have reduced the number of non-target bands. Produced internal primers may have also helped reduce nonspecific amplicon production (Ford and Rose 1995). Other gradients involving varying PCR ingredients could have been used as well, an example of this would be a magnesium gradient.

Although there was not a lot of variation between genes when designing the primers there can still be variation between the commercial strains used for the PCR optimisation. This is because rhizobia can undergo transconjugation (Broughton, Samrey et al. 1987). This means that even though they can be the same biovar they can transfer plasmids to other bacteria they meet causing them to mutate quickly. This can explain why there can be differences in desiccation tolerance within and between the collection sites. This could be a reason why WSM1325 was continuously producing multiple bands and often went unamplified. The commercial strain WSM1689 may also be significantly different from WSM1325 that was used for PCR optimisation. The gels do indicate there is a difference between the commercial strains. WSM1325 produced multiple bands on most of the gels. This may be due to the genetic differences between the commercial strain, WSM1325 being a commercial strain of subclover and cc275e and TA1 white clover.

Absence of amplification in some isolates for some strains may indicate the target gene was not present for example; isolate 41 may not contain the genes *tre*Y and *tre*Z. It is possible that the genes are presents/ absence may vary between the isolates which could be why some of isolates do not have bands present. However, it is more likely that it was experimental error. Rhizobia most likely contain the genes they will be expressed at different levels between bacteria and located in differing positions and locations in the bacterial genome as per the contigs figure 10, 11 and appendixes 22, 23, and 24.

Alignment of the nine sequences for the trehalose-6-phosphate phosphatase gene *otsB* showed that there was variable between strains for *ots*B. However, the only correction was that may indicate differences in tolerance was the first amino acid substitution which changed polarity between isolates 42, 31(polar amino acid, Serine) and the remaining isolates (non- polar amino acid, Alanine) as per table 8. This substitution may change the secondary structure of the protein.

The phylogeny tree does correlate with chapter 2 biofilm or PEG data. The isolates selected for tolerance in the biofilm are grouped semi grouped, except for isolate 42 being with isolate 32. Alternatively, three of the non-tolerant isolates were together, 47, 50 and 51. Work done by McIntyre, Davies et al. (2007) suggest that the accumulation of trehalose maybe controlled at the post transcriptional level. As the previous chapter suggested there is no correlation from collection site. With the majority of the amino acid polymorphisms being conservative it is unlikely that there is an alteration in protein function. The commercial strains were not desiccation tolerant. There was little polymorphic variation between the reference commercial genome genes used for primer design and the isolate that produced clean sequences. The alignment of the isolates contained a nucleotide identity of 96.6%. When the reference strains were combined the identity dropped down to 96.11%. Table 3.5 shows some amino acid polymorphisms between isolates.

3.5 Conclusion

In summary all five of the genes were present in the selected isolates. The PCR products did not produce readable sequences. The only one of the five genes trehalose-6-phosphate phosphatase (*ots*B) was successfully sequenced for 9 out of 10 isolates. This chapter has shown that there is genetic variation of the *ots*B gene. The *ots*B gene variation showed some differences that could help group isolates in the future. However, to get the complete picture the entire sequences needed to be mapped, including the promoter regions to see if there is variation in regulation.

Chapter 4 – Comparison of the survival of *Rhizobium leguminosarum* strains selected for desiccation tolerance in soil

4.1 Introduction

In vitro tests are useful to identify traits in strains of bacteria that may be related to their ability to withstand dry. However, desiccation stress is an extreme stress that evokes a complex signalling cascade in bacterial cells (Zhu 2002, Apel and Hirt 2004). This makes it difficult to be sure that the response observed in the laboratory will be replicated in the field. However, the soil is a complex substrate. The ability of soil bacteria to tolerate drying will be affected by numerous biotic and abiotic factors. Several examples of biotic factor effecting rhizobia survival are presence absence of host legume (Mary, Ochin et al. 1985), presence of predators (Pena-Cabriales and Alexander 1979). Nutrients and pH have a significant influence on saprophyte survival (Osa-Afiana and Alexander 1982). Various farming practices will also influence soil bacteria survival towards drying, such as cultivation, which is used to reduce diseases in some crops (MacLeod, Vanstone et al.
2008).

The aim of this chapter was to determine whether strains that produced large amounts of polysaccharide and that were tolerant of PEG could survive better in dry soil. This was achieved by i) developing erythromycin resistant mutants of the bacterial strains for use in unsterilized soil, ii) monitoring the CFU of the bacteria over time in a soil that had been subject to drying.

4.2 Methodology

4.2.1 Development of Erythromycin resistant strains

Two isolates were selected that appeared to produce high biofilm (isolate 53) and tolerate
PEG (isolate 42) from the chapter 2 and compared to nontolerant isolates (isolates 47 and 50). TA1 was included as the current commercial strain which is not desiccation tolerant. The isolates were grown on YMA plates for 3 days at 25^{o}C in the dark. After 3 days a single colony was plated onto YMA containing erythromycin (10 ppm). The process of plating on antibiotic amended YMA media following by “resting” on unmended media was replicated with gradually

increasing concentrations of erythromycin of 30 ppm, 50 ppm, 70 ppm, 100 ppm, and 125 ppm and erythromycin (10ppm). Plates were incubated on each medium for 3 days before resting or increasing the antibiotic concentration. When the isolates were able to grow on 125 ppm antibiotic concentration the durability of the erythromycin resistance was checked by alternating on unamended and amended media for 3 times each. The genotype of the resultant erythromycin resistance colony was then determined and compared to that of the initial colony.

4.2.2 Colony ERIC-PCR

Colony PCR was done by touching a bacterial colony lightly with a sterile pipette and then touching to the master mix prepared as described in section 2.2.2.3. ERIC-PCR and gel electrophoresis was done as described in section 2.2.2.4.

4.2.3 Conformation of ability to grow in PEG

Each strain and its corresponding erythromycin mutant were grown in 60% polyethylene glycol as described in section 2.2.3.2. For each strain triplicates tubes were prepared each containing a 3 mL of a 1 x 10^7 CFU/mL culture in 60% PEG solution in a 50 mL tube with 30 mL 60% PEG solution and placed in an incubator at 28^{o}C and 210 rpm for 0, 6, 24, and 56 hours respectively. At each time point the optical density was measured at 600nm using a spectrophotometer. Statistical analysis was done as described in section 2.2.3.2

4.2.4 Growth in soil

Soil from the Winchmore Research Station was selected because it had no detectable background rhizobia present (Ridgway pers. comm.). Three lots of 20 g of soil were weighed out and dried in a 100^{o}C soil oven for 12 h to achieve a soil moisture content of approximately 0%. When the soil has dried it was put into an empty Petri dish and inoculated with a bacterial culture of strain TA1 to achieve 10^8 CFU/g soil (prepared as described in section 2.2.3.1). The soil moisture content was raised to 20% by adding 4 mL of inoculant. Each inoculant was mixed thoroughly with the soil. For dilution plating 2 g soil was placed into a 15 mL tube containing 9 mL of 0.001% triton X in water and shaken in a wrist shaker for 20 min (10^{-1}). This mixture was serially diluted a further 6 times (10^{-2}-10^{-7}). One hundred uL of each dilution was spread plated onto YMA medium amended with 125 ppm erythromycin and 0.1% benomyl (to inhibit fungal growth) in triplicate.

Plates were incubated for 7 d at 28°C in the dark and then the number of colonies on the plate counted and recorded. The inoculated dry soil was incubated at room temperature for 3 d and the process of plating repeated.

4.2.4.2 Desiccation tolerance in soil

The ability of each of the five antibiotic mutants (42, 47, 50, 53 and TA1) to survive in dry soil and a negative control to which YMB was added without bacteria. Approximately 500 g of Winchmore soil was dried in an oven at 100^{o}C for 12 h and then sieved under sterile conditions with a 700 m sieve. Twenty grams of soil was weighed into three Petri plates for each isolate (42, 47, 50, 53, TA1) and a negative control (18 dishes total). Each Petri dish was inoculated with 10^{8} CFU/g of soil and brought up to 20% soil moisture capacity (4 mL of inoculant). Two grams of soil was taken as a time zero (T=0) sample and dilution plated as described in section 4.2.4.1. The plates were then placed in a randomised complete block design, covered with tinfoil and incubated at room temperature. Two grams of soil was taken from each of the 18 Petri dishes at 4, 9, 14, 21 and 45 d after inoculation. Not all 7 dilutions were plated as the pilot study had identified the likely concentration at each incubation time. The dilutions for each incubation period are listed in the table 4.2.

Dilution plate on 125ppm erythromycin and 1% Benomyl.

Table 4.1 This table illustrates what dilutions were plated for each incubation time.

Incubation days	Dilutions 2	3	4	5	6
0		•	•	•	•
4		•	•	•	
9	•	•	•		
14	•	•	•		
21	•	•	•		
42	•	•	•		

4.2.4.3 Statistical Analysis

The growth in PEG over time was analysed as described in section 2.2.3.2.3.

The CFU counts over time were analysed using ANOVA as per section 2.2.3.2.3

4.3 Results

4.3.1 Colony ERIC-PCR

Isolate42 and 53 (putative dry tolerant), 47 and 50 (not dry tolerant) were selected for the soil assays. The colony ERIC-PCR was used to confirm that the genotypes remain the same between the antibiotic mutant and non-mutant isolates (data not shown).

4.3.2 Tolerance to 60% PEG

There were no significant differences between the growth of strains at 0 (P=0.384), 6 (P=0.995), 24 (P=0.088) or 52 (P=0.810) h incubation.

4.3.2 Soil persistence

All the antibiotic mutants produced colonies following the dilution plating of inoculated soil. There was no significant difference between isolates at T= 0, 9, 14 or 21 d after inoculation (LSD≤0.05) (Table 4.2). There were significant differences between strains at 4 and 45 d after inoculation. After 4 d incubation strain 42 had a greater number of CFUs than isolate 47, but was not different from strains 50, 53 and TA1. At 45 days after inoculation strain 53 had a greater number of CFU compared to strain 50 and TA1 but was not different from strain 47 and 42. For all strains there was a decrease in CFU over time (Figure 4.2 and Table 4.1).

Table 4.2 Statistically significant groups of antibiotic isolate mutants.

	Mean log of cells/ml on days after inoculation into soil (cells per ml)					
Isolate	Day 0	Day 4	Day 9	Day 14	Day 21	Day 45
42	8.24 (1.4×10^9)	5.18a (1.4×10^7)	5.29 (1.7×10^7)	4.92 (8.3×10^6)	4.87 (7.5×10^6)	4.35ab (2.4×10^6)
47	8.52 (2.0×10^9)	3.43b (2.3×10^5)	5.00 (9.8×10^6)	4.14 (1.5×10^6)	4.39 (2.7×10^6)	4.13abc (1.4×10^6)
50	8.11 (1.2×10^9)	4.88ab (7.7×10^6)	5.31 (1.8×10^7)	3.47 (2.5×10^5)	4.70 (5.3×10^6)	3.09c (7.9×10^4)
53	7.61 (6.5×10^8)	4.92ab (8.3×10^6)	5.18 (1.4×10^7)	3.61 (3.8×10^5)	5.11 (1.2×10^7)	4.83a (6.9×10^6)
TA1	7.69 (7.2×10^8)	4.86ab (7.4×10^6)	5.09 (1.2×10^7)	3.44 (2.3×10^5)	4.86 (7.4×10^6)	3.27bc (1.4×10^5)
LSD	1.312	1.484	0.725	1.666	0.877	1.226

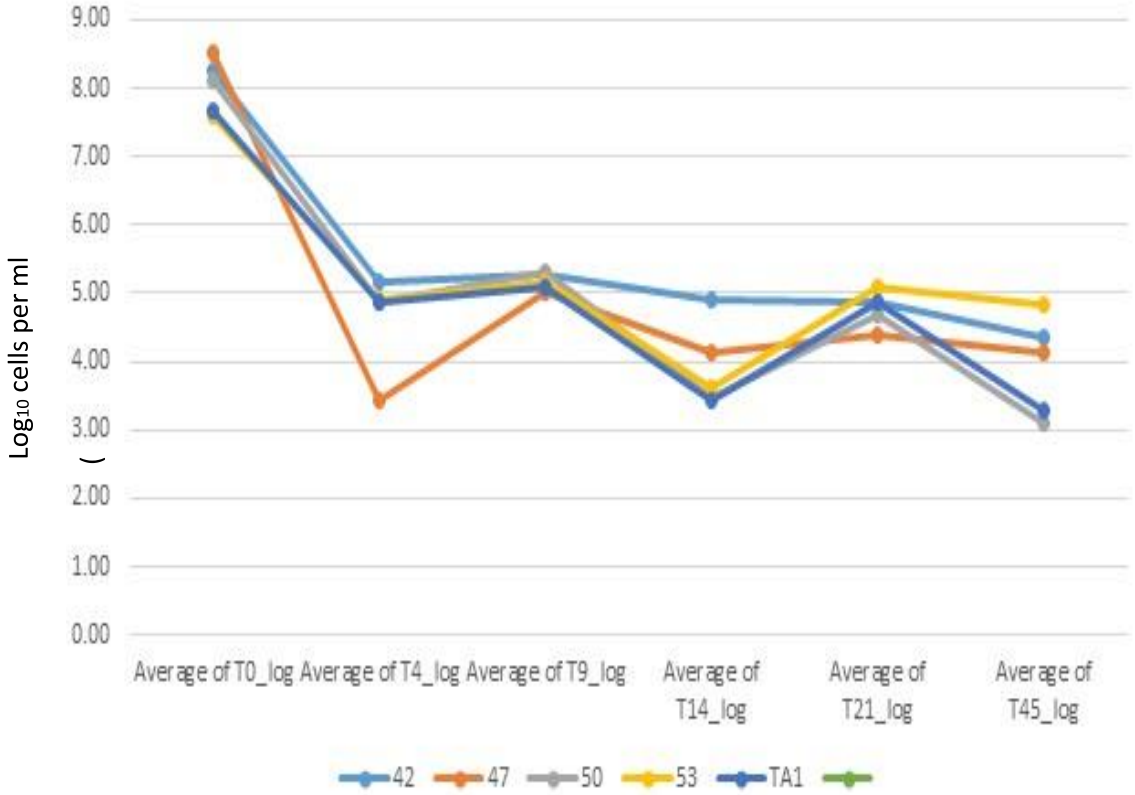


Figure 4.1 Log transformation of the relative survival (CFU) in soil of *R. leguminosarum* strains over time.

4.4 Discussion

This chapter attempted to confirm that strains of *R. leguminosarum* able to produce a large biofilm and to grow well in high concentrations of PEG could survive better in dry soil. Two isolates were chosen based on their performance in the assays of chapter 2 and these were isolate 53 (best biofilm producer) and 42 (best tolerance to 60% PEG). However, issues arise when trying to evaluate bacteria in soil due to the high background of bacteria able to grow on agar medium. Antibiotic mutants have been used as a simple and effective way to monitor bacteria in complex ecosystems (Compeau, Al-Achi et al. 1988). It is possible that the process of generating the antibiotic mutants may affect some aspects of the rhizobia such as the nodulation efficiency as has been shown previously (Pankhurst 1977, Lewis,
Bromfield et al. 1987). However, they are a useful way to monitor selected isolates in soil (Juhnke, Mathre et al. 1987). The results showed they were a useful way to reduce contamination during dilution plating part and allowed the target colonies to be counted. Although there was some growth on control plates they were distinguishable from the inoculated rhizobia.

The PEG assay was used to confirm that the ranking of the strains remained the same instead of the biofilm because antibiotics can influence biofilm production (Stewart and Costerton 2001). The biofilms are altered because it is a survival mechanism bacterium use to survive against antibiotics (Stewart and Costerton 2001). The results showed that there was no difference between the antibiotic mutants and non-mutants in the PEG assay. This meant that the dry tolerance of strain 42 to PEG could not be confirmed. There was also no difference between the mutant and wild type strains indicating that the creation of antibiotic mutants but as there was no longer any difference between high and low biofilm producers it was unclear whether this meant that the production of the mutants had affect desiccation tolerance. To get a more accurate result more work should be done to compare antibiotic mutants to the wild type to determine if the process has affected rhizobia phenotypes. Such assays could include most probably number to assess if the symbiotic potential is altered. A biofilm assay to determine if there are alterations in excreted polysaccharides.

When introduced into soil and placed under a desiccation pressure the results showed variability between strains. This showed that the two tolerant (42 and

53) strains survived well in dry soil but were not different at all-time points from all other strains. Variability could have been decreased by more than three replicates, which may have been insufficient for soil which is a complex and variable substrate. The results supported the hypothesis that selection of strains using the PEG and biofilm assay could identify more dry strains. Tolerance to drying may have arisen due to a bacterium's ability for adaptive mutation (Foster 2005) which allows them to adapt to differing environments (Bjedov, Tenaillon et al. 2003). The isolates were chosen on the basis that the isolates from the dry sites with low SMD would be desiccation tolerant and vice versa. The annual SMD of the soil from which the strains were derived did not correlate well with the survival in dry soil. This is evident by isolate 53 being characterised as desiccation tolerant. Rhizobia are known to have variation between drought and other stress tolerances (Munns and Keyser 1981, Subbarao, Johansen et al. 1990). However, the extent and biological understanding of this is not well known (Osa-Afiana and Alexander 1982). The soil persistence in this experiment gave a good indication that the strains do exhibit different levels of tolerance over a 45 day drying period. To gain a more realistic assessment of field survival this experiment needs to be done on a larger scale with a plant host present and in conjunction with a fast-drying period (Bushby and Marshall 1977, Mary, Ochin et al. 1985, Vriezen, De Bruijn et al. 2007).

Isolate 42 and 53 were selected based on their ability to form a biofilm and tolerance to PEG. Their good performance in the drying soil suggests that these are useful assays for identifying strains with ability to tolerate desiccation. Isolate 42 had the highest absorbance in the PEG assay and isolate 53 the largest biofilms, which is known to have a significant effect on survival in strong abiotic stresses (Grover, Ali et al. 2011)**.** This result may also have implications for the commercial coating of clover seed with *R. leguminosarum*. Seeds coated with TA1 have been shown to have low survival rates on seed and in soil (LOWTHER[1] and Kerr 2011). This meant that it was a good reference to compare to since it is known to have poor survivability in dry conditions.

Although the soil assay suggested that isolates 42 and 53 could survive well in dry soil, other factors can also be attributed to survival in the soil. Both abiotic and biotic factors can influence the survival (Pena-Cabriales and Alexander 1979). An

example of this is the presence or absence of host legume. Rhizobia are not found in high concentration in the absence of its host legume plant (Mary, Ochin et al. 1985). Experiments to assess how plants can affect survival in dry soil would be difficult as low soil moisture is likely to lead to plant death. A wild type unsterilized soil was used for the persistence experiment described here. This had very low concentration of rhizobia and drying the soil at 100°C should have further depleted the microbes present in the soil, including small eukaryotic organisms i.e. protozoa that prey upon rhizobia (Pena-Cabriales and Alexander 1979). Thus, similar work in more soil types should be used to extend this work as this also contributes to survival rates when subject to a drought stress (Pena-Cabriales and Alexander 1979). In future manipulating drying speeds could also give a better indication of desiccation tolerance (Mary, Ochin et al. 1985).

4.5 Conclusion

In conclusion the antibiotic mutants were an effective means to identify rhizobia in a soil system. However, the strains did not differ when grown in PEG so the relative raking of the strains and mutants could not be confirmed. The experimental results suggested that the two strains selected for dry tolerance were better than TA1 at surviving in dry soil. However, the experimental design needs to be improved further to account for other variables that could be influencing survival of rhizobia strains when subjected to desiccation stresses.

5 Concluding Discussion

5.1.1 Novel findings

The overall aim of this thesis was to determine if there was any variation between strains of *R. leguminosarum* in desiccation tolerances. This was tested by:

1. Characterising desiccation tolerance *in vitro* using two bioassays
2. Investigating genetic variation between strains for genes involved in the production of the polysaccharide trehalose
3. Confirming desiccation tolerance of selected strains via persistence in drying soil.

The present study showed there was genetic diversity of *R. leguminosarum* strains recovered from sites of contrasting SMD in New Zealand (Seehaver 2014, De Meyer, De Beuf et al. 2015). This was illustrated by the genotyping done in chapter 2. This study also showed that *R. leguminosarum* strains exhibited variation in their ability to produce biofilms and to grow in high concentrations of PEG tolerance. This was expected because the *R. leguminosarum* strains were genetically diverse and variability in response to different stresses is well known in bacteria.

It was hypothesized that isolates grown in sites that had a high SMD would exhibit a higher tolerance to desiccation stress compared to isolates collected from sites that had a low SMD. However, the assays used in this study showed significant between isolates, but this did not relate well with the SMD of the soil of origin. These results signify that in most soils there will be strains with variation in desiccation tolerant regardless of the SMD. The use of bioassays for biofilm production and PEG tolerance provided a good starting point for characterising and ranking desiccation tolerance in *R. leguminosarum*. However, these assays need improving to increase their accuracy, perhaps by increasing the replication, in the future.

Desiccation evokes a complex array of genetic signalling in bacteria (Cooper 2007, Kobayashi and Broughton 2008). Since trehalose a significant influence on organism survival when exposed to desiccation it was a good starting point to

look for genetic variation in genes associated with desiccation tolerance. The primers develop for the 5 trehalose biosynthetic genes (*otsA*, *otsB, tre*Y, *tre*Z and *tre*S) (Engelhard 2004, Cytryn, Sangurdekar et al. 2007, Reina-Bueno, Argandoña et al. 2012), all produced amplicons of the target size. However only the primers designed otsB produced DNA sequences that were clean and analysable. There were single nucleotide polymorphisms that resulted in amino acid substitutions. The variation between amino acids did correlate back to the ranking slightly as strains that were selected for tolerance grouped into different groups than those with no evidence of tolerance.

It is difficult to account for all variables in when studying bacteria in soil that can influence *R. leguminosarum* (Pena-Cabriales and Alexander 1979). The use of antibiotic mutants proved to be a useful way to analyse desiccation tolerance in soil. There were no significant differences between the mutants and wildtype strains. The persistence in soil confirmed that strains 42 and 53 were more tolerant and other strains of dry soil conditions.

The work described here provides a good starting point for selecting *R. leguminosarum* for desiccation tolerance as it showed that:

1. There is variation between strains in their ability to grow in PEG amended media and to produce biofilms
2. Strains with the highest biofilm and greatest growth in PEG amended media were able to survive in soil better than other strains
3. *In vitro* assays are useful to select strains
4. There is variation between strains in genes associated with desiccation tolerance

5.1.2 Future work

This chapter produced results that allowed strains to be ranked on their desiccation tolerance. However, there were technical issues with some of the methodology:

- If this research were to be repeated a larger sample size (more strains) should be used. This would potentially give a more accurate understanding of whether desiccation tolerance could achieve using a

rapid screening process based on molecular markers such as ERIC-PCR fingerprint. For example, isolate 42 and 41 were the only isolates found with genotypes J and I (figure 2.3). Both of these were quite desiccation tolerant. If a larger population was studied, it is possible that the genotype may correlate with evidence for tolerance.

- The biofilm assay has technical issues as the crystal violet is a non-specific stain (Fujishige, Kapadia et al. 2006, Burton, Yakandawala et al. 2007). It stains any polysaccharide that is bound to the surface of the wells. If this experiment was repeated another stain should be used that is specific for trehalose. The protocol
  also needs to be modified as the rinsing process often removed some of the biofilm (Fujishige, Kapadia et al. 2006). This assay was also done using strains in nonstressed conditions. Better results could be obtained if it was repeated with the bacterial cells in a stressed state (eg. with PEG).
- The PEG assay also provided difficult to use as its viscosity made it difficult to mix and pipette into cuvettes for the absorbance readings. Also, in chapter 1 the PEG assay was done in duplicates and should be repeated with a higher number of replicates to account for system variability.

To gain greater understanding of what might happen under field conditions a third assay should be introduced. Rhizobia are rhizosphere colonizers and therefore the host plant is likely to interact with the isolates to help them survive desiccation. One experiment that could determine this is a most probable number (MPN) (Kremer and Peterson 1983). This would determine different rates of nodulation under dry conditions. The rhizosphere may influence survival. *R. leguminosarum* by altering their carbon source during desiccation or by providing a refuge. Further work on carbon use and its relationship to desiccation tolerance could be explored using a □□ MicroResp analysis (Wigley (2017).

When the genes for the trehalose pathway were amplified all *R. leguminosarum* strains produced a band of the target size. However, only the *ots*B produce useful sequences. In the future the other four need to be sequenced to determine if they have any variation. More primers could be developed so the entire genes can be amplified, including the upstream, regulatory regions. As the contigs

(figure 3.1, 3.2 and appendixes 3.1, 3.2 and 3.3) illustrates there is variations in gene content and order this may also influence gene activation and expression. (Wassarman 2002, McIntyre, Davies et al. 2007) found that trehalose is controlled at the post transcriptional level and that could also be interesting to investigate further.

Soil persistence was successful experiment to confirm desiccation tolerance. If this experiment was to be done again a second treatment should be added to it as the rate of drying can influence desiccation tolerance (Mary, Ochin et al. 1985).

5.2 Final conclusion

The experimental procedures produced a simple way to identify if R. *leguminosarum* strains exhibit desiccation tolerance. There were differences between some strains that allowed the identification of desiccation tolerant isolates. There appeared to be no significant differences regarding desiccation tolerance and the SMD of soils the strains originated from. It supports the possibility that better commercial *R. leguminosarum* strains could be selected for survival in desiccation prone habitats. Any future work should look at a broader range of sites and a larger number of isolates. Additional bioassays could be implemented to make the ranking more accurate and to provide greater simulation of field conditions.

References

http://www.premierbiosoft.com/tech_notes/PCR_Primer_Design.html

https://www.bioinformatics.org/sms/iupac.html

Apel, K. and H. Hirt (2004). "Reactive oxygen species: metabolism, oxidative stress, and signal transduction." Annu. Rev. Plant Biol. **55**: 373-399.

Argüelles, J. C. (2000). "Physiological roles of trehalose in bacteria and yeasts: a comparative analysis." Archives of Microbiology **174**(4): 217-224.

Avonce, N., et al. (2006). "Insights on the evolution of trehalose biosynthesis." BMC evolutionary biology **6**(1): 109.

Bardin, S., et al. (1996). "A phosphate transport system is required for symbiotic nitrogen fixation by Rhizobium meliloti." Journal of bacteriology **178**(15): 4540-4547.

Bayoumi, T., et al. (2008). "Application of physiological and biochemical indices as a screening technique for drought tolerance in wheat genotypes." African journal of biotechnology **7**(14).

Becana, M. and J. I. Sprent (1987). "Nitrogen fixation and nitrate reduction in the root nodules of legumes." Physiologia Plantarum **70**(4): 757-765.

Billi, D., et al. (2001). "Gene transfer to the desiccation-tolerant cyanobacteriumChroococcidiopsis." Journal of bacteriology **183**(7): 2298-2305.

Billi, D. and M. Potts (2002). "Life and death of dried prokaryotes." Research in microbiology **153**(1): 7-12.

Bjedov, I., et al. (2003). "Stress-induced mutagenesis in bacteria." Science **300**(5624): 14041409.

Blum, A. and A. Ebercon (1981). "Cell membrane stability as a measure of drought and heat tolerance in wheat1." Crop Science **21**(1): 43-47.

Bobik, C., et al. (2006). "FixJ: a major regulator of the oxygen limitation response and late symbiotic functions of Sinorhizobium meliloti." Journal of bacteriology **188**(13): 4890-4902.

Bomchil, N., et al. (2003). "Identification and characterization of a Vibrio cholerae gene, mbaA, involved in maintenance of biofilm architecture." Journal of bacteriology **185**(4): 1384-1390.

Broughton, W., et al. (1987). "Ecological genetics of Rhizobium meliloti: symbiotic plasmid transfer in the Medicago sativa rhizosphere." FEMS Microbiology Letters **40**(2-3): 251-255.

Burton, E., et al. (2007). "A microplate spectrofluorometric assay for bacterial biofilms." Journal of industrial microbiology & biotechnology **34**(1): 1-4.

Buscaill, P. and S. Rivas (2014). "Transcriptional control of plant defence responses." Current opinion in plant biology **20**: 35-46.

Bushby, H. and K. Marshall (1977). "Some factors affecting the survival of root-nodule bacteria on desiccation." Soil Biology and Biochemistry **9**(3): 143-147.

Bushby, H. V. A. and K. C. Marshall (1977). "Some factors affecting the survival of rootnodule bacteria on desiccation." Soil Biology and Biochemistry **9**(3): 143-147.

Chang, W.-S., et al. (2007). "Alginate production by Pseudomonas putida creates a hydrated microenvironment and contributes to biofilm architecture and stress tolerance under water-limiting conditions." Journal of bacteriology **189**(22): 8290-8299.

Coffey, B. M. and G. G. Anderson (2014). Biofilm formation in the 96-well microtiter plate. Pseudomonas Methods and Protocols, Springer**:** 631-641.
Compeau, G., et al. (1988). "Survival of rifampin-resistant mutants of Pseudomonas fluorescens and Pseudomonas putida in soil systems." Applied and environmental microbiology **54**(10): 2432-2438.

Cooper, J. (2007). "Early interactions between legumes and rhizobia: disclosing complexity in a molecular dialogue." Journal of applied microbiology **103**(5): 1355-1365.

Cubero, J. and J. Graham (2002). "Genetic relationship among worldwide strains of Xanthomonas causing canker in citrus species and design of new primers for their identification by PCR." Applied and environmental microbiology **68**(3): 1257-1264.

Cytryn, E. J., et al. (2007). "Transcriptional and physiological responses of Bradyrhizobium japonicum to desiccation-induced stress." Journal of bacteriology **189**(19): 6751-6762.

De Ley, J. and A. Rassel (1965). "DNA base composition, flagellation and taxonomy of the genus Rhizobium." Journal of general microbiology **41**(1): 85-92.

De Meyer, S. E., et al. (2015). "A large diversity of non-rhizobial endophytes found in legume root nodules in Flanders (Belgium)." Soil Biology and Biochemistry **83**: 1-11.

De Smet, K. A., et al. (2000). "Three pathways for trehalose biosynthesis in mycobacteria." Microbiology **146**(1): 199-208.

Deaker, R., et al. (2004). "Legume seed inoculation technology—a review." Soil Biology and Biochemistry **36**(8): 1275-1288.

Denison, R. F. and E. T. Kiers (2011). "Life histories of symbiotic rhizobia and mycorrhizal fungi." Current Biology **21**(18): R775-R785.

Dieffenbach, C., et al. (1993). "General concepts for PCR primer design." PCR methods appl **3**(3): S30-S37.

Downie, J. A. (2010). "The roles of extracellular proteins, polysaccharides and signals in the interactions of rhizobia with legume roots." FEMS microbiology reviews **34**(2): 150-170.

Duodu, S., et al. (2009). "Competitiveness of a native Rhizobium leguminosarum biovar trifolii strain for nodule occupancy is manifested during infection." Plant and soil **318**(1-2): 117.

Eckert, K. A. and T. A. Kunkel (1991). "DNA polymerase fidelity and the polymerase chain reaction." Genome Research **1**(1): 17-24.

Engelhard, M. (2004). Trehalose and the nitrogen fixing nodule symbiosis of legumes:
studies on rhizobia deficient in the trehalose-6-phosphate synthase gene" ots" A, University of Basel.

Faulkner, C. and S. Robatzek (2012). "Plants and pathogens: putting infection strategies and defence mechanisms on the map." Current opinion in plant biology **15**(6): 699-707.

Fernandez, O., et al. (2010). "Trehalose and plant stress responses: friend or foe?" Trends in plant science **15**(7): 409-417.

Ford, O. and E. A. Rose (1995). "Long-PCR: Long-distance PCR." PCR Primer: A Laboratory Manual. Cold Spring Harbor Laboratory Press, New York: 63-77.

Foster, P. L. (2005). "Stress responses and genetic variation in bacteria." Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis **569**(1): 3-11.
Fujishige, N. A., et al. (2006). "Investigations of Rhizobium biofilm formation." FEMS microbiology ecology **56**(2): 195-206.

Gaunt, M., et al. (2001). "Phylogenies of atpD and recA support the small subunit rRNAbased classification of rhizobia." International journal of systematic and evolutionary microbiology **51**(6): 2037-2048.

Geetha, S. and S. J. Joshi (2013). "Engineering rhizobial bioinoculants: a strategy to improve iron nutrition." The Scientific World Journal **2013**.

Giller, K., et al. (2016). The legume-rhizobia symbiosis and assessing the need to inoculate. Working with rhizobia, Australasian Society for the Study of Animal Behaviuor**:** 15-24.

González, J. E. and M. M. Marketon (2003). "Quorum sensing in nitrogen-fixing rhizobia." Microbiology and Molecular Biology Reviews **67**(4): 574-592.

Grover, M., et al. (2011). "Role of microorganisms in adaptation of agriculture crops to abiotic stresses." World Journal of Microbiology and Biotechnology **27**(5): 1231-1240.

Hargrove, M. S., et al. (2000). "Crystal structure of a nonsymbiotic plant hemoglobin." Structure **8**(9): 1005-1014.

Hebb, M., et al. (1998). "PCR as an ecological tool to determine the establishment and persistence of Rhizobium strains introduced into the field as seed inoculant Diane." Australian Journal of Agricultural Research **49**(6): 923-934.

Hirsch, A. M. (1992). "Developmental biology of legume nodulation." New Phytologist **122**(2): 211-237.

Howieson, J. and R. Ballard (2004). "Optimising the legume symbiosis in stressful and competitive environments within southern Australia—some contemporary thoughts." Soil Biology and Biochemistry **36**(8): 1261-1273.

Hückelhoven, R. (2007). "Transport and secretion in plant–microbe interactions." Current opinion in plant biology **10**(6): 573-579.

Humann, J. L., et al. (2009). "Regulatory and DNA repair genes contribute to the desiccation resistance of Sinorhizobium meliloti Rm1021." Applied and environmental microbiology **75**(2): 446-453.

Innis, M. and D. Gelfand (1999). "Optimization of PCR: conversations between Michael and David." PCR applications: protocols for functional genomics: 3-22.

Issa, S., et al. (1993). "Active movement of chickpea and bean rhizobia in dry soil." Soil Biology and Biochemistry **25**(7): 951-958.

Jabado, O. J., et al. (2006). "Greene SCPrimer: a rapid comprehensive tool for designing degenerate primers from multiple sequence alignments." Nucleic acids research **34**(22): 6605-6611.

Jensen, M., et al. (1993). "Rapid identification of bacteria on the basis of polymerase chain reaction-amplified ribosomal DNA spacer polymorphisms." Applied and environmental microbiology **59**(4): 945-952.

Juhnke, M. E., et al. (1987). "Identification and characterization of rhizosphere-competent bacteria of wheat." Applied and environmental microbiology **53**(12): 2793-2799.

Kalendar, R., et al. (2009). "FastPCR software for PCR primer and probe design and repeat search." Genes, Genomes and Genomics **3**(1): 1-14.

Kobayashi, H. and W. Broughton (2008). Fine-tuning of symbiotic genes in rhizobia: flavonoid signal transduction cascade. Nitrogen-fixing leguminous symbioses, Springer**:** 117-152.

Kosek, M., et al. (2012). "Facilitated molecular typing of Shigella isolates using ERIC-PCR." The American journal of tropical medicine and hygiene **86**(6): 1018-1025.

Kremer, R. J. and H. L. Peterson (1983). "Effects of carrier and temperature on survival of Rhizobium spp. in legume inocula: development of an improved type of inoculant." Applied and environmental microbiology **45**(6): 1790-1794.

Krusell, L., et al. (2005). "The sulfate transporter SST1 is crucial for symbiotic nitrogen fixation in Lotus japonicus root nodules." The Plant Cell Online **17**(5): 1625-1636.

Kwok, S., et al. (1994). "A guide to the design and use of mismatched and degenerate." Genome Res **3**: S39-S47.

Lang, M. and V. Orgogozo (2011). "Identification of homologous gene sequences by PCR with degenerate primers." Molecular Methods for Evolutionary Genetics: 245-256.

Leung, K. T., et al. (2004). "A comparison of AFLP and ERIC-PCR analyses for discriminating Escherichia coli from cattle, pig and human sources." FEMS microbiology ecology **47**(1): 111-119.

Lewis, D., et al. (1987). "Effect of rifampin resistance on nodulating competitiveness of Rhizobium meliloti." Canadian journal of microbiology **33**(4): 343-345.

Liu, Y., et al. (2005). "Transcriptome analysis of Shewanella oneidensis MR-1 in response to elevated salt conditions." Journal of bacteriology **187**(7): 2501-2507.

LOWTHER[1], W. and G. Kerr (2011). "White clover seed inoculation and coating in New Zealand."

MacLeod, B., et al. (2008). "Root disease under intensive cereal production systems."

Mahajan, S. and N. Tuteja (2005). "Cold, salinity and drought stresses: an overview." Archives of biochemistry and biophysics **444**(2): 139-158.

Martínez-Romero, E. and J. Caballero-Mellado (1996). "Rhizobium phylogenies and bacterial genetic diversity." Critical Reviews in Plant Sciences **15**(2): 113-140.

Martinez-Romero, E. and M. Rosenblueth (1990). "Increased bean (Phaseolus vulgaris L.) nodulation competitiveness of genetically modified Rhizobium strains." Applied and environmental microbiology **56**(8): 2384-2388.

Maruta, K., et al. (1996). "Cloning and sequencing of trehalose biosynthesis genes from Arthrobacter sp. Q36." Biochimica et Biophysica Acta (BBA)-General Subjects **1289**(1): 1013.

Mary, P., et al. (1985). "Rates of drying and survival of Rhizobium meliloti strains during storage at different relative humidities." Applied and environmental microbiology **50**(2): 207-211.

Matthews, P., et al. (1999). "Livestock farming systems in New Zealand." & JGH White (Ed.), New Zealand Pasture and Crop Science: 101-116.
Maunoury, N., et al. (2008). Cell biology of nodule infection and development. Nitrogenfixing Leguminous Symbioses, Springer**:** 153-189.

McDougall S (2015) Generating a “tool-box for investigating drought tolerance in *Rhizobium leguminosarum*. Summer studentship reports 16pp; Lincoln University, Christchurch, New Zealand.

McIntyre, H. J., et al. (2007). "Trehalose biosynthesis in Rhizobium leguminosarum bv. trifolii and its role in desiccation tolerance." Applied and environmental microbiology **73**(12): 3984-3992.

McIntyre, H. J., et al. (2007). "Trehalose biosynthesis in Rhizobium leguminosarum bv. trifolii and its role in desiccation tolerance." Applied and environmental microbiology **73**(12): 3984-3992.

Menna, P., et al. (2006). "Molecular phylogeny based on the 16S rRNA gene of elite rhizobial strains used in Brazilian commercial inoculants." Systematic and Applied Microbiology **29**(4): 315-332.

Mergaert, P., et al. (2006). "Eukaryotic control on bacterial cell cycle and differentiation in the Rhizobium–legume symbiosis." Proceedings of the National Academy of Sciences of the United States of America **103**(13): 5230-5235.

Mills, A., et al. (2015). "'MaxClover'grazing experiment: I. Annual yields, botanical composition and growth rates of six dryland pastures over nine years." Grass and forage science **70**(4): 557-570.

Munns, D. and H. Keyser (1981). "Response of Rhizobium strains to acid and aluminium stress." Soil Biology and Biochemistry **13**(2): 115-118.

Nangul, A., et al. (2013). "Nodule occupancy by Rhizobium leguminosarum strain WSM1325 following inoculation of four annual Trifolium species in Canterbury, New Zealand." New Zealand Journal of Agricultural Research **56**(3): 215-223.

O'Hara, G. (2001). "Nutritional constraints on root nodule bacteria affecting symbiotic nitrogen fixation: a review." Animal Production Science **41**(3): 417-433.

O'Hara, G. (2001). "Nutritional constraints on root nodule bacteria affecting symbiotic nitrogen fixation: a review." Australian Journal of Experimental Agriculture **41**(3): 417-433.

O'Hara, G., et al. (2016). Taxonomy and physiology of rhizobia, Australian Centre for International Agricultural Research.

Ökmen, B. and G. Doehlemann (2014). "Inside plant: biotrophic strategies to modulate host immunity and metabolism." Current opinion in plant biology **20**: 19-25.

Oldroyd, G. E., et al. (2005). "Peace talks and trade deals. Keys to long-term harmony in legume-microbe symbioses." Plant Physiology **137**(4): 1205-1210.

Osa-Afiana, L. and M. Alexander (1982). "Differences among cowpea rhizobia in tolerance to high temperature and desiccation in soil." Applied and environmental microbiology **43**(2): 435-439.

Pankhurst, C. E. (1977). "Symbiotic effectiveness of antibiotic-resistant mutants of fast-and slow-growing strains of Rhizobium nodulating Lotus species." Canadian journal of microbiology **23**(8): 1026-1033.

Pena-Cabriales, J. and M. Alexander (1979). "Survival of Rhizobium in Soils Undergoing Drying 1." Soil Science Society of America Journal **43**(5): 962-966.

Potts, M. (1994). "Desiccation tolerance of prokaryotes." Microbiological reviews **58**(4): 755-805.
Potts, M. (1994). "Desiccation tolerance of prokaryotes." Microbiological reviews **58**(4): 755.

Potts, M. (1999). "Mechanisms of desiccation tolerance in cyanobacteria." European Journal of Phycology **34**(04): 319-328.

Potts, M. (2001). "Desiccation tolerance: a simple process?" TRENDS in Microbiology **9**(11): 553-559.

Radhouane, L. (2007). "Response of Tunisian autochthonous pearl millet (Pennisetum glaucum (L.) R. Br.) to drought stress induced by polyethylene glycol (PEG) 6000." African journal of biotechnology **6**(9).

Rawsthorne, S., et al. (1980). "Carbon and nitrogen metabolism in legume root nodules." Phytochemistry **19**(3): 341-355.

Reina-Bueno, M., et al. (2012). "Role of trehalose in heat and desiccation tolerance in the soil bacterium Rhizobium etli." BMC microbiology **12**(1): 207.

Russo, D. M., et al. (2006). "Proteins exported via the PrsD-PrsE type I secretion system and the acidic exopolysaccharide are involved in biofilm formation by Rhizobium leguminosarum." Journal of bacteriology **188**(12): 4474-4486.

Sandhya, V., et al. (2009). "Alleviation of drought stress effects in sunflower seedlings by the exopolysaccharides producing Pseudomonas putida strain GAP-P45." Biology and fertility of soils **46**(1): 17-26.

Seehaver, S. A. (2014). "Effect of Inoculation on Cover Crop Legume Nodule Occupation and Associated Rhizobia Diversity."

Shokralla, S., et al. (2012). "Next-generation sequencing technologies for environmental DNA research." Molecular ecology **21**(8): 1794-1805.

Simon-Rosin, U., et al. (2003). "Molecular and cellular characterisation of LjAMT2; 1, an ammonium transporter from the model legume Lotus japonicus." Plant molecular biology **51**(1): 99-108.

Sipos, R., et al. (2007). "Effect of primer mismatch, annealing temperature and PCR cycle number on 16S rRNA gene-targetting bacterial community analysis." FEMS microbiology ecology **60**(2): 341-350.

Sprent, J. I. (2009). "Global Distribution of Legumes." Legume Nodulation: A Global Perspective: 35-50.

Stacey, G. (2007). The Rhizobium-legume nitrogen-fixing symbiosis. Biology of the Nitrogen Cycle, Elsevier London**:** 147.

Steinweg, J. M., et al. (2013). "Microbial responses to multi-factor climate change: effects on soil enzymes." Frontiers in microbiology **4**.

Stewart, P. S. and J. W. Costerton (2001). "Antibiotic resistance of bacteria in biofilms." The lancet **358**(9276): 135-138.

Suárez, R., et al. (2008). "Improvement of drought tolerance and grain yield in common bean by overexpressing trehalose-6-phosphate synthase in rhizobia." Molecular PlantMicrobe Interactions **21**(7): 958-966.

Subbarao, G., et al. (1990). "Response of the pigeonpea-Rhizobium symbiosis to salinity stress: variation among Rhizobium strains in symbiotic ability." Biology and fertility of soils **9**(1): 49-53.

Suzaki, T. and M. Kawaguchi (2014). "Root nodulation: a developmental program involving cell fate conversion triggered by symbiotic bacterial infection." Current opinion in plant biology **21**: 16-22.

Udvardi, M. K. and D. A. Day (1997). "Metabolite transport across symbiotic membranes of legume nodules." Annual review of plant biology **48**(1): 493-523.

Van den Berg, L. and Y. Zeng (2006). "Response of South African indigenous grass species to drought stress induced by polyethylene glycol (PEG) 6000." South African Journal of Botany **72**(2): 284-286.

van Ham, R., et al. (2016). "Soil moisture deficit selects for desiccation tolerant Rhizobium leguminosarum bv. trifolii." Applied Soil Ecology **108**: 371-380.

Van Wees, S., et al. (2008). "Plant immune responses triggered by beneficial microbes." Current opinion in plant biology **11**(4): 443-448.

Vance, C. and G. Heichel (1991). "Carbon in N2 fixation: limitation or exquisite adaptation." Annual review of plant biology **42**(1): 373-390.

Vance, C. P. and J. S. Gantt (1992). "Control of nitrogen and carbon metabolism in root nodules." Physiologia Plantarum **85**(2): 266-274.

Vanderlinde, E. M., et al. (2010). "Identification of a novel ABC transporter required for desiccation tolerance, and biofilm formation in Rhizobium leguminosarum bv. viciae 3841." FEMS microbiology ecology **71**(3): 327-340.

Versalovic, J., et al. (1991). "Distribution of repetitive DNA sequences in eubacteria and application to finerpriting of bacterial enomes." Nucleic acids research **19**(24): 6823-6831.

Versalovic, J., et al. (1994). "Genomic fingerprinting of bacteria using repetitive sequencebased polymerase chain reaction." Methods in molecular and cellular biology **5**(1): 25-40.

Vlassak, K. M., et al. (1997). "Factors influencing nodule occupancy by inoculant rhizobia." Critical Reviews in Plant Sciences **16**(2): 163-229.

Vriezen, J. A., et al. (2007). "Responses of rhizobia to desiccation in relation to osmotic stress, oxygen, and temperature." Applied and environmental microbiology **73**(11): 34513459.

Vriezen, J. A. C., et al. (2007). "Responses of rhizobia to desiccation in relation to osmotic stress, oxygen, and temperature." Applied and environmental microbiology **73**(11): 34513459.

Wang, D., et al. (2012). "Symbiosis specificity in the legume–rhizobial mutualism." Cellular microbiology **14**(3): 334-342.

Wassarman, K. M. (2002). "Small RNAs in bacteria: diverse regulators of gene expression in response to environmental changes." Cell **109**(2): 141-144.

Weisburg, W. G., et al. (1991). "16S ribosomal DNA amplification for phylogenetic study." Journal of bacteriology **173**(2): 697-703.
Wigley, K. (2017). Aspects of soil ecology of rhizobia affecting nodule occupancy on lucerne and white clover, Lincoln University.

Zahran, H. H. (1999). "Rhizobium-legume symbiosis and nitrogen fixation under severe conditions and in an arid climate." Microbiology and molecular biology reviews **63**(4): 968989.

Zepeda Mendoza, M. L., et al. (2015). "Environmental genes and genomes: understanding the differences and challenges in the approaches and software for their analyses." Briefings in bioinformatics **16**(5): 745-758.

Zhu, J.-K. (2002). "Salt and drought stress signal transduction in plants." Annual review of plant biology **53**(1): 247-273.

Zribi, K., et al. (2004). "Distribution and genetic diversity of rhizobia nodulating natural populations of Medicago truncatula in Tunisian soils." Soil Biology and Biochemistry **36**(6):
903-908.

Appendix

A 2 Chapter 2

A 1 Recipes A1.1 Yeast Mannitol Agar

1 g yeast extract
4 g mannitol
0.5 g dipotassium phosphate
0.2 g magnesium sulphate
0.1 g sodium chloride
15 g agar
1 l water
Autoclave for 15 minutes at 121°C and 15 Psi

A 1.2 Yeast Mannitol Broth

1 g yeast extract
4 g mannitol
0.5 g dipotassium phosphate
0.2 g magnesium sulphate
0.1 g sodium chloride
1 l water

Autoclave for 15 minutes at 121°C and 15 Psi

A2 Chapter 2 A 1.3 16s partial sequence

Isolate 31
GACGGGTGAGTAACGCGTGGGAATCTACCCTTGACTACGGAATAACGCAGGGAAA
CTTGTGC
TAATACCGTATGTGTCCTTCGGGAGAAAGATTTATCGGTCAAGGATGAGCCCGCG
TTGGATT
AGCTAGTTGGTGGGGTAAAGGCCTACCAAGGCGACGATCCATAGCTGGTCTGAGA
GGATGAT
CAGCCACATTGGGACTGAGACACGGCCCAAACTCCTACGGGAGGCAGCAGTGGGG
AATATTG
GACAATGGGCGCAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCCTAGG
GTTGTAA
AGCTCTTTCACCGGAGAAGATAATGACGGTATCCGGAGAAGAAGCCCCGGCTAAC
TTCGTGC
CAGCAGCCGCGGTAATACGAAGGGGGCTAGCGTTGTTCGGAATTACTGGGCGTAA
AGCGCAC
GTAGGCGGATCGATCAGTCAGGGGTGAAATCCCAGGGCTCAACCCTGGAACTGCC
TTTGATA
CTGTCGATCTGGAGTATGGAAGAGGTGAGTGGAATTCCGAGTGTAGAGGTGAAAT
TCGTAGA
TATTCGGAGGAACACCAGTGGCGAAGGCGGCTCACTGGTCCATTACTGACGCTGA
GGTGCGA AAGCGTGGGG

Isolate 32
GAGCGGCAGACGGGTGAGTAACGCGTGGGAATCTACCCTTGACTACGGAATAACG
CAGGGAA

ACTTGTGCTAATACCGTATGTGTCCTTCGGGAGAAAGATTTATCGGTCAAGGATG
AGCCCGC
GTTGGATTAGCTAGTTGGTGGGGTAAAGGCCTACCAAGGCGACGATCCATAGCTG
GTCTGAG
AGGATGATCAGCCACATTGGGACTGAGACACGGCCCAAACTCCTACGGGAGGCAG
CAGTGGG
GAATATTGGACAATGGGCGCAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAG
GCCCTAG
GGTTGTAAAGCTCTTTCACCGGAGAAGATAATGACGGTATCCGGAGAAGAAGCCC
CGGCTAA
CTTCGTGCCAGCAGCCGCGGTAATACGAAGGGGGCTAGCGTTGTTCGGAATTACT
GGGCGTA
AAGCGCACGTAGGCGGATCGATCAGTCAGGGGTGAAATCCCAGGGCTCAACCCTG
GAACTGC
CTTTGATACTGTCGATCTGGAGTATGGAAGAGGTGAGTGGAATTCCGAGTGTAGA
GGTGAAA
TTCGTAGATATTCGGAGGAACACCAGTGGCGAAGGCGGCTCACTGGTCCATTACT
GACGCTG
AGGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTA
GTCCAC

Isolate 33
TGCAGTCGAGCGCCCGCAGGGAGCGGCAGACGGGTGAGTAACGCGTGGGAATCTA
CCCTTGA
CTACGGAATAACGCAGGGAAACTTGTGCTAATACCGTATGTGTCCTTCGGGAGAA
AGATTTA
TCGGTCAAGGATGAGCCCGCGTTGGATTAGCTAGTTGGTGGGGTAAAGGCCTACC
AAGGCGA
CGATCCATAGCTGGTCTGAGAGGATGATCAGCCACATTGGGACTGAGACACGGCC
CAAACTC
CTACGGGAGGCAGCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCC
ATGCCGC
GTGAGTGATGAAGGCCCTAGGGTTGTAAAGCTCTTTCACCGGAGAAGATAATGAC
GGTATCC
GGAGAAGAAGCCCCGGCTAACTTCGTGCCAGCAGCCGCGGTAATACGAAGGGGGC
TAGCGTT
GTTCGGAATTACTGGGCGTAAAGCGCACGTAGGCGGATCGATCAGTCAGGGGTGA
AATCCCA
GGGCTCAACCCTGGAACTGCCTTTGATACTGTCGATCTGGAGTATGGAAGAGGTG
AGTGGAA
TTCCGAGTGTAGAGGTGAAATTCGTAGATATTCGGAGGAACACCAGTGGCGAAGG
CGGCTCA
CTGGTCCATTACTGACGCTGAGGTGCGAAAGCGTGGGGAGCAAACAGGATTAGAT
ACCCTGG
TAGTCCACGCCGTAAACGATGAATGTTAGCCGTCGGGCAGTATACTGTTCGGTGG
CGCAGCT AAACGCA

Isolate 34
CCGCAGGAGCGGCAGACGGGTGAGTAACGCGTGGGAATCTACCCTTGACTACGGA
ATAACGC
AGGGAAACTTGTGCTAATACCGTATGTGTCCTTCGGGAGAAAGATTTATCGGTCA
AGGATGA

GCCCGCGTTGGATTAGCTAGTTGGTGGGGTAAAGGCCTACCAAGGCGACGATCCA
TAGCTGG
TCTGAGAGGATGATCAGCCACATTGGGACTGAGACACGGCCCAAACTCCTACGGG
AGGCAGC
AGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCCATGCCGCGTGAGTG
ATGAAGG
CCCTAGGGTTGTAAAGCTCTTTCACCGGAGAAGATAATGACGGTATCCGGAGAAG
AAGCCCC
GGCTAACTTCGTGCCAGCAGCCGCGGTAATACGAAGGGGGCTAGCGTTGTTCGGA
ATTACTG
GGCGTAAAGCGCACGTAGGCGGATCGATCAGTCAGGGGTGAAATCCCAGGGCTCA
ACCCTGG
AACTGCCTTTGATACTGTCGATCTGGAGTATGGAAGAGGTGAGTGGAATTCCGAG
TGTAGAG
GTGAAATTCGTAGATATTCGGAGGAACACCAGTGGCGAAGGCGGCTCACTGGTCC
ATTACTG
ACGCTGAGGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCA
CGCCGTA
AACGATGAATGTTAGCCGTCGGGCAGTATACTGTTCGGTGGCGCAGCTAACGCAT
TAA

Isolate 35
TGCAGTCGAGCGCCCCGCAGGGGAGCGGCAGACGGGTGAGTAACGCGTGGGAATC
TACCCTT
GACTACGGAATAACGCAGGGAAACTTGTGCTAATACCGTATGTGTCCTTCGGGAG
AAAGATT
TATCGGTCAAGGATGAGCCCGCGTTGGATTAGCTAGTTGGTGGGGTAAAGGCCTA
CCAAGGC
GACGATCCATAGCTGGTCTGAGAGGATGATCAGCCACATTGGGACTGAGACACGG
CCCAAAC
TCCTACGGGAGGCAGCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAG
CCATGCC
GCGTGAGTGATGAAGGCCCTAGGGTTGTAAAGCTCTTTCACCGGAGAAGATAATG
ACGGTAT
CCGGAGAAGAAGCCCCGGCTAACTTCGTGCCAGCAGCCGCGGTAATACGAAGGGG
GCTAGCG
TTGTTCGGAATTACTGGGCGTAAAGCGCACGTAGGCGGATCGATCAGTCAGGGGT
GAAATCC
CAGGGCTCAACCCTGGAACTGCCTTTGATACTGTCGATCTGGAGTATGGAAGAGG
TGAGTGG
AATTCCGAGTGTAGAGGTGAAATTCGTAGATATTCGGAGGAACACCAGTGGCGAA
GGCGGCT
CACTGGTCCATTACTGACGCTGAGGTGCGAAAGCGTGGGGAGCAAACAGGATTAG
ATACCCT
GGTAGTCCACGCCGTAAACGATGAATGTTAGCCGTCGGGCAGTATACTGTTC

Isolate 36
TCGAGCGCCCCGCAGGGGAGCGGCAGACGGGTGAGTAACGCGTGGGAATCTACCC
TTGACTA
CGGAATAACGCAGGGAAACTTGTGCTAATACCGTATGTGTCCTTCGGGAGAAAGA
TTTATCG

GTCAAGGATGAGCCCGCGTTGGATTAGCTAGTTGGTGGGGTAAAGGCCTACCAAG
GCGACGA
TCCATAGCTGGTCTGAGAGGATGATCAGCCACATTGGGACTGAGACACGGCCCAA
ACTCCTA
CGGGAGGCAGCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCCATG
CCGCGTG
AGTGATGAAGGCCCTAGGGTTGTAAAGCTCTTTCACCGGAGAAGATAATGACGGT
ATCCGGA
GAAGAAGCCCCGGCTAACTTCGTGCCAGCAGCCGCGGTAATACGAAGGGGGCTAG
CGTTGTT
CGGAATTACTGGGCGTAAAGCGCACGTAGGCGGATCGATCAGTCAGGGGTGAAAT
CCCAGGG
CTCAACCCTGGAACTGCCTTTGATACTGTCGATCTGGAGTATGGAAGAGGTGAGT
GGAATTC
CGAGTGTAGAGGTGAAATTCGTAGATATTCGGAGGAACACCAGTGGCGAAGGCGG
CTCACTG
GTCCATTACTGACGCTGAGGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACC
CTGGTAG
TCCACGCCGTAAACGATGAATGTTAGCCGTCGGGCAGTATAC
TGTTCG
Isolate 37
ACACATGCAGTCGAGCGCCCGCAGGGAGCGGCAGACGGGTGAGTAACGCGTGGGA
ATCTACC
CTTGACTACGGAATAACGCAGGGAAACTTGTGCTAATACCGTATGTGTCCTTCGG
GAGAAAG
ATTTATCGGTCAAGGATGAGCCCGCGTTGGATTAGCTAGTTGGTGGGGTAAAGGC
CTACCAA
GGCGACGATCCATAGCTGGTCTGAGAGGATGATCAGCCACATTGGGACTGAGACA
CGGCCCA
AACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATC
CAGCCAT
GCCGCGTGAGTGATGAAGGCCCTAGGGTTGTAAAGCTCTTTCACCGGAGAAGATA
ATGACGG
TATCCGGAGAAGAAGCCCCGGCTAACTTCGTGCCAGCAGCCGCGGTAATACGAAG
GGGGCTA
GCGTTGTTCGGAATTACTGGGCGTAAAGCGCACGTAGGCGGATCGATCAGTCAGG
GGTGAAA
TCCCAGGGCTCAACCCTGGAACTGCCTTTGATACTGTCGATCTGGAGTATGGAAG
AGGTGAG
TGGAATTCCGAGTGTAGAGGTGAAATTCGTAGATA

Isolate 38
CGAGCGCCCGCAGGGGAGCGGCAGACGGGTGAGTAACGCGTGGGAATCTACCCTT
GACTACG
GAATAACGCAGGGAAACTTGTGCTAATACCGTATGTGTCCTTCGGGAGAAAGATT
TATCGGT
CAAGGATGAGCCCGCGTTGGATTAGCTAGTTGGTGGGGTAAAGGCCTACCAAGGC
GACGATC
CATAGCTGGTCTGAGAGGATGATCAGCCACATTGGGACTGAGACACGGCCCAAAC
TCCTACG
GGAGGCAGCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCCATGCC
GCGTGAG

TGATGAAGGCCCTAGGGTTGTAAAGCTCTTTCACCGGAGAAGATAATGACGGTAT
CCGGAGA
AGAAGCCCCGGCTAACTTCGTGCCAGCAGCCGCGGTAATACGAAGGGGGCTAGCG
TTGTTCG
GAATTACTGGGCGTAAAGCGCACGTAGGCGGATCGATCAGTCAGGGGTGAAATCC
CAGGGCT
CAACCCTGGAACTGCCTTTGATACTGTCGATCTGGAGTATGGAAGAGGTGAGTGG
AATTCCG
AGTGTAGAGGTGAAATTCGTAGATATTCGGAGGAACACCAGTGGCGAAGGCGGCT
CACTGGT
CCATTACTGACGCTGAGGTGCGAAAGCGTG

Isolate 39
TGCAGTCGAGCGCCCGCAGGGAGCGGCAGACGGGTGAGTAACGCGTGGGAATCTA
CCCTTGA
CTACGGAATAACGCAGGGAAACTTGTGCTAATACCGTATGTGTCCTTCGGGAGAA
AGATTTA
TCGGTCAAGGATGAGCCCGCGTTGGATTAGCTAGTTGGTGGGGTAAAGGCCTACC
AAGGCGA
CGATCCATAGCTGGTCTGAGAGGATGATCAGCCACATTGGGACTGAGACACGGCC
CAAACTC
CTACGGGAGGCAGCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCC
ATGCCGC
GTGAGTGATGAAGGCCCTAGGGTTGTAAAGCTCTTTCACCGGAGAAGATAATGAC
GGTATCC
GGAGAAGAAGCCCCGGCTAACTTCGTGCCAGCAGCCGCGGTAATACGAAGGGGGC
TAGCGTT
GTTCGGAATTACTGGGCGTAAAGCGCACGTAGGCGGATCGATCAGTCAGGGGTGA
AATCCCA
GGGCTCAACCCTGGAACTGCCTTTGATACTGTCGATCTGGAGTATGGAAGAGGTG
AGTGGAA
TTCCGAGTGTAGAGGTGAAATTCGTAGATATTCGGAGGAACACCAGTGGCGAAGG
CGGCTCA
CTGGTCCATTACTGACGCTGAGGTGCGAAAGCGTGGGGGAGC

Isolate 40
GCAGTCGAGCGCCCCGCAGGGGAGCGGCAGACGGGTGAGTAACGCGTGGGAATCT
ACCCTTG
ACTACGGAATAACGCAGGGAAACTTGTGCTAATACCGTATGTGTCCTTCGGGAGA
AAGATTT
ATCGGTCAAGGATGAGCCCGCGTTGGATTAGCTAGTTGGTGGGGTAAAGGCCTAC
CAAGGCG
ACGATCCATAGCTGGTCTGAGAGGATGATCAGCCACATTGGGACTGAGACACGGC
CCAAACT
CCTACGGGAGGCAGCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGC
CATGCCG
CGTGAGTGATGAAGGCCCTAGGGTTGTAAAGCTCTTTCACCGGAGAAGATAATGA
CGGTATC
CGGAGAAGAAGCCCCGGCTAACTTCGTGCCAGCAGCCGCGGTAATACGAAAGGGG
GCTAGCG
TTTGTTTCGGAATTACTGGGCGTAAAGCGCACGTAGGCGGATCGATCAGTCAGGG
GTGAAAT

CCCAGGGCTCAACCCTGGAACTGCCTTTGATACTGTCGATCTGGAGTATGGAAGA
GGTGAGT
GGAATTCCGAGTGTAGAGGTGAAATTCGTAGATATTCGGAGGAACACCAGTGGCG
AAGGCGG
CTCACTGGTCCATTACTGACGCTGAGGTGCGAAA

Isolate 41
CCATGCAGTCGAGCGCCCCGCAAGGGGAGCGGCAGACGGGTGAGTAACGCGTGGG
AATCTAC
CCTTGACTACGGAATAACGCAGGGAAACTTGTGCTAATACCGTATGTGTCCTTCG
GGAGAAA
GATTTATCGGTCAAGGATGAGCCCGCGTTGGATTAGCTAGTTGGTGGGGTAAAGG
CCTACCA
AGGCGACGATCCATAGCTGGTCTGAGAGGATGATCAGCCACATTGGGACTGAGAC
ACGGCCC
AAACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGAT
CCAGCCA
TGCCGCGTGAGTGATGAAGGCCCTAGGGTTGTAAAGCTCTTTCACCGGAGAAGAT
AATGACG
GTATCCGGAGAAGAAGCCCCGGCTAACTTCGTGCCAGCAGCCGCGGTAATACGAA
GGGGGCT
AGCGTTGTTCGGAATTACTGGGCGTAAAGCGCACGTAGGCGGATCGATCAGTCAG
GGGTGAA
ATCCCAGGGCTCAACCCTGGAACTGCCTTTGATACTGTCGATCTGGAGTATGGAA
GAGGTGA
GTGGAATTCCGAGTGTAGAGGTGAAATTCGTAGATATTCGGAGGAACACCAGTGG
CGAAGGC
GGCTCACTGGTCCATTACTGACGCTGAGGTGCGAAAGCGTGGGGGAGCAAACAGG
ATTAGAT
ACCCTGGGTAGTCCACGCCGTAACGATGAATGTTAGCCGTCGGGCAGTATA

Isolate 42
GTACCATGCAGTCGAGCGCCCCGCAGGGGGAGCGGCAGACGGGTGAGTAACGCGT
GGGAATC
TACCCTTGACTACGGAATAACGCAGGGAAACTTGTGCTAATACCGTATGTGTCCT
TCGGGAG
AAAGATTTATCGGTCAAGGATGAGCCCGCGTTGGATTAGCTAGTTGGTGGGGTAA
AGGCCTA
CCAAGGCGACGATCCATAGCTGGTCTGAGAGGATGATCAGCCACATTGGGACTGA
GACACGG
CCCAAACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAATGGGCGCAAGCCT
GATCCAG
CCATGCCGCGTGAGTGATGAAGGCCCTAGGGTTGTAAAGCTCTTTCACCGGAGAA
GATAATG
ACGGTATCCGGAGAAGAAGCCCCGGCTAACTTCGTGCCAGCAGCCGCGGTAATAC
GAAGGGG
GCTAGCGTTGTTCGGAATTACTGGGCGTAAAGCGCACGTAGGCGGATCGATCAGT
CAGGGGT
GAAATCCCAGGGCTCAACCCTGGAACTGCCTTTGATACTGTCGATCTGGAGTATG
GAAGAGG
TGAGTGGAATTCCGAGTGTAGAGGTGAAATTCGTAGATATTCGGGAGGAACACCA
GTGGCGA

AGGCGGCTCACTGGTCCATTACTGACGCTGAGGTGCGAAAG
CGTGGG
Isolate 43
GGGTGAGTAACGCGTGGGAATCTACCCTTGACTACGGAATAACGCAGGGAAACTT
GTGCTAA
TACCGTATGTGTCCTTCGGGAGAAAGATTTATCGGTCAAGGATGAGCCCGCGTTG
GATTAGC
TAGTTGGTGGGGTAAAGGCCTACCAAGGCGACGATCCATAGCTGGTCTGAGAGGA
TGATCAG
CCACATTGGGACTGAGACACGGCCCAAACTCCTACGGGAGGCAGCAGTGGGGAAT
ATTGGAC
AATGGGCGCAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCCTAGGGTT
GTAAAGC
TCTTTCACCGGAGAAGATAATGACGGTATCCGGAGAAGAAGCCCCGGCTAACTTC
GTGCCAG
CAGCCGCGGTAATACGAAGGGGGCTAGCGTTGTTCGGAATTACTGGGCGTAAAGC
GCACGTA
GGCGGATCGATCAGTCAGGGGTGAAATCCCAGGGCTCAACCCTGGAACTGCCTTT
GATACTG
TCGATCTGGAGTATGGAAGAGGTGAGTGGAATTCCGAGTGTAGAGGTGAAATTCG
TAGATAT
TCGGAGGAACACCAGTGGCGAAGGCGGCTCACTGGTCCATTACTGACGCTGAGGT
GCGAAAG
CGTGGGGAGCAAACAGGATTAGATACCCTGGT

Isolate 44
ACCATGCAGTCGAGCGCCCGCAGGGAGCGGCAGACGGGTGAGTAACGCGTGGGAA
TCTACCC
TTGACTACGGAATAACGCAGGGAAACTTGTGCTAATACCGTATGTGTCCTTCGGG
AGAAAGA
TTTATCGGTCAAGGATGAGCCCGCGTTGGATTAGCTAGTTGGTGGGGTAAAGGCC
TACCAAG
GCGACGATCCATAGCTGGTCTGAGAGGATGATCAGCCACATTGGGACTGAGACAC
GGCCCAA
ACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCC
AGCCATG
CCGCGTGAGTGATGAAGGCCCTAGGGTTGTAAAGCTCTTTCACCGGAGAAGATAA
TGACGGT
ATCCGGAGAAGAAGCCCCGGCTAACTTCGTGCCAGCAGCCGCGGTAATACGAAGG
GGGCTAG
CGTTGTTCGGAATTACTGGGCGTAAAGCGCACGTAGGCGGATCGATCAGTCAGGG
GTGAAAT
CCCAGGGCTCAACCCTGGAACTGCCTTTGATACTGTCGATCTGGAGTATGGAAGA
GGTGAGT
GGAATTCCGAGTGTAGAGGTGAAATTCGTAGATATTCGGAGGAACACCAGTGGCG
AAGGCGG
CTCACTGGTCCATTACTGACGCTGAGGTGCGAAAGCGTGGGGAGCAAACAGGATT
AGATACC
CTGGTAGTCCACGCCGTAAACGATGAATGTTAGCCGTCGGGCAGTATACTGT
Isolate 45
TGCAGTCGAGCGCCCGCAGGAGCGGCAGACGGGTGAGTAACGCGTGGGAATCTAC
CCTTGAC

TACGGAATAACGCAGGGAAACTTGTGCTAATACCGTATGTGTCCTTCGGGAGAAA
GATTTAT
CGGTCAAGGATGAGCCCGCGTTGGATTAGCTAGTTGGTGGGGTAAAGGCCTACCA
AGGCGAC
GATCCATAGCTGGTCTGAGAGGATGATCAGCCACATTGGGACTGAGACACGGCCC
AAACTCC
TACGGGAGGCAGCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCCA
TGCCGCG
TGAGTGATGAAGGCCCTAGGGTTGTAAAGCTCTTTCACCGGAGAAGATAATGACG
GTATCCG
GAGAAGAAGCCCCGGCTAACTTCGTGCCAGCAGCCGCGGTAATACGAAGGGGGCT
AGCGTTG
TTCGGAATTACTGGGCGTAAAGCGCACGTAGGCGGATCGATCAGTCAGGGGTGAA
ATCCCAG
GGCTCAACCCTGGAACTGCCTTTGATACTGTCGATCTGGAGTATGGAAGAGGTGA
GTGGAAT
TCCGAGTGTAGAGGTGAAATTCGTAGATATTCGGAGGAACACCAGTGGCGAAGGC
GGCTCAC
TGGTCCATTACTGACGCTGAGGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATA
CCCTGGT
AGTCCACGCCGTAAACGATGAATGTTAGCCGTCGGGCAGTATACT
GTTCGG

Isolate 46

TGCAGTCGAGCGCCCGCAGGGAGCGGCAGACGGGTGAGTAACGCGTGGGAATCTA
CCCTTGA
CTACGGAATAACGCAGGGAAACTTGTGCTAATACCGTATGTGTCCTTCGGGAGAA
AGATTTA
TCGGTCAAGGATGAGCCCGCGTTGGATTAGCTAGTTGGTGGGGTAAAGGCCTACC
AAGGCGA
CGATCCATAGCTGGTCTGAGAGGATGATCAGCCACATTGGGACTGAGACACGGCC
CAAACTC
CTACGGGAGGCAGCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCC
ATGCCGC
GTGAGTGATGAAGGCCCTAGGGTTGTAAAGCTCTTTCACCGGAGAAGATAATGAC
GGTATCC
GGAGAAGAAGCCCCGGCTAACTTCGTGCCAGCAGCCGCGGTAATACGAAGGGGGC
TAGCGTT
GTTCGGAATTACTGGGCGTAAAGCGCACGTAGGCGGATCGATCAGTCAGGGGTGA
AATCCCA
GGGCTCAACCCTGGAACTGCCTTTGATACTGTCGATCTGGAGTATGGAAGAGGTG
AGTGGAA
TTCCGAGTGTAGAGGTGAAATTCGTAGATATTCGGAGGAACACCAGTGGCGAAGG
CGGCTCA
CTGGTCCATTACTGACGCTGAGGTGCGAAAGCGTGGGGAG

Isolate 47

ATGCAGTCGAGCGCCCCGCAAGGGGAGCGGCAGACGGGTGAGTAACGCGTGGGAA
TCTACCC
TTGACTACGGAATAACGCAGGGAAACTTGTGCTAATACCGTATGTGTCCTTCGGG
AGAAAGA
TTTATCGGTCAAGGATGAGCCCGCGTTGGATTAGCTAGTTGGTGGGGTAAAGGCC
TACCAAG

GCGACGATCCATAGCTGGTCTGAGAGGATGATCAGCCACATTGGGACTGAGACAC
GGCCCAA
ACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCC
AGCCATG
CCGCGTGAGTGATGAAGGCCCTAGGGTTGTAAAGCTCTTTCACCGGAGAAGATAA
TGACGGT
ATCCGGAAAAAAAAGCCCCGGCTAACTTCGTGCCAGCAGCCGCGGTAATACGAAG
GGGGCTA
GCGTTGTTCGGAATTACTGGGCGTAAAGCGCACGTAGGCCGGATCGATCAGTCGG
GGGTGAA
ATCCCAGGGCTCAACCCTGGAACTGCCTTTTGATACTGTCGATCTGGAGTATGGA
AGAGGTG
AGTGGAATTCCGAGTGTAGAGGTGAAAATTCGTAAAATTTTCGGAAGAAACACCA
GTGGCGA
AGGGGGCTCACTGGTCCATTACTGACGCTGGAGGTGCGAAAAGCGTGGGGGAGCA
AACAGGA
TTAGAATCCCCTGGTAGTCCACGCCGTAAACGATGAATGGTTAGCCGTCGGGCAA
GTATACT
GTTCGGTGGCGCAGCTAACGCATTAAACATTTCGCCTGGGGAGTACGGTCGCAAG
ATTAAAA CTCAAGGAA

Isolate 48
TGCAGTCGAGCGCCCGCAGGGGAGCGGCAGACGGGTGAGTAACGCGTGGGAATCT
ACCCTTG
ACTACGGAATAACGCAGGGAAACTTGTGCTAATACCGTATGTGTCCTTCGGGAGA
AAGATTT
ATCGGTCAAGGATGAGCCCGCGTTGGATTAGCTAGTTGGTGGGGTAAAGGCCTAC
CAAGGCG
ACGATCCATAGCTGGTCTGAGAGGATGATCAGCCACATTGGGACTGAGACACGGC
CCAAACT
CCTACGGGAGGCAGCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGC
CATGCCG
CGTGAGTGATGAAGGCCCTAGGGTTGTAAAGCTCTTTCACCGGAGAAGATAATGA
CGGTATC
CGGAGAAGAAGCCCCGGCTAACTTCGTGCCAGCAGCCGCGGTAATACGAAGGGGG
CTAGCGT
TGTTCGGAATTACTGGGCGTAAAGCGCACGTAGGCGGATCGATCAGTCAGGGGTG
AAATCCC
AGGGCTCAACCCTGGAACTGCCTTTGATACTGTCGATCTGGAGTATGGAAGAGGT
GAGTGGA
ATTCCGAGTGTAGAGGTGAAATTCGTAGATATTCGGAGGAACACCAGTGGCGAAG
GCGGCTC
ACTGGTCCATTACTGACGCTGAGGTGCGAAAGCGTGGGGA
Isolate 49
GGGAGCGGCAGACGGGTGAGTAACGCGTGGGAATCTACCCTTGACTACGGAATAA
CGCAGGG
AAACTTGTGCTAATACCGTATGTGTCCTTCGGGAGAAAGATTTATCGGTCAAGGA
TGAGCCC
GCGTTGGATTAGCTAGTTGGTGGGGTAAAGGCCTACCAAGGCGACGATCCATAGC
TGGTCTG
AGAGGATGATCAGCCACATTGGGACTGAGACACGGCCCAAACTCATACGGGAGGC
AGCAGTG

GGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCCATGCCGCGTGAGTGATGA
AGGCCCT
AGGGTTGTAAAGCTCTTTCACCGGAGAAGATAATGACGGTATCCGGAGAAGAAGC
CCCGGCT
AACTTCGTGCCAGCAGCCGCGGTAATACGAAGGGGGCTAGCGTTGTTCGGAATTA
CTGGGCG
TAAAGCGCACGTAGGCGGATCGATCAGTCAGGGGTGAAATCCCAGGGCTCAACCC
TGGAACT
GCCTTTGATACTGTCGATCTGGAGTATGGAAGAGGTGAGTGGAATTCCGAGTGTA
GAGGTGA
AATTCGTAGATATTCGGAGGAACACCAGTGGCGAAGGCGGCTCACTGGTCCATTA
CTGACGC
TGAGGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCC
TGGTAG
Isolate 50
CGAGCGCCCGCAGGGAGCGGCAGACGGGTGAGTAACGCGTGGGAATCTACCCTTG
ACTACGG
AATAACGCAGGGAAACTTGTGCTAATACCGTATGTGTCCTTCGGGAGAAAGATTT
ATCGGTC
AAGGATGAGCCCGCGTTGGATTAGCTAGTTGGTGGGGTAAAGGCCTACCAAGGCG
ACGATCC
ATAGCTGGTCTGAGAGGATGATCAGCCACATTGGGACTGAGACACGGCCCAAACT
CATACGG
GAGGCAGCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCCATGCCG
CGTGAGT
GATGAAGGCCCTAGGGTTGTAAAGCTCTTTCACCGGAGAAGATAATGACGGTATC
CGGAGAA
GAAGCCCCGGCTAACTTCGTGCCAGCAGCCGCGGTAATACGAAGGGGGCTAGCGT
TGTTCGG
AATTACTGGGCGTAAAGCGCACGTAGGCGGATCGATCAGTCAGGGGTGAAATCCC
AGGGCTC
AACCCTGGAACTGCCTTTGATACTGTCGATCTGGAGTATGGAAGAGGTGAGTGGA
ATTCCGA
GTGTAGAGGTGAAATTCGTAGATATTCGGAGGAACACCAGTGGCGAAGGCGGCTC
ACTGGTC
CATTACTGACGCTGAGGTGCGAAAGCGTGGGGAGCAAACAGG

Isolate 51
ATGCAGTCGAGCGCCCGCAGGGAGCGGCAGACGGGTGAGTAACGCGTGGGAATCT
ACCCTTG
ACTACGGAATAACGCAGGGAAACTTGTGCTAATACCGTATGTGTCCTTCGGGAGA
AAGATTT
ATCGGTCAAGGATGAGCCCGCGTTGGATTAGCTAGTTGGTGGGGTAAAGGCCTAC
CAAGGCG
ACGATCCATAGCTGGTCTGAGAGGATGATCAGCCACATTGGGACTGAGACACGGC
CCAAACT
CCTACGGGAGGCAGCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGC
CATGCCG
CGTGAGTGATGAAGGCCCTAGGGTTGTAAAGCTCTTTCACCGGAGAAGATAATGA
CGGTATC
CGGAGAAGAAGCCCCGGCTAACTTCGTGCCAGCAGCCGCGGTAATACGAAGGGGG
CTAGCGT

TGTTCGGAATTACTGGGCGTAAAGCGCACGTAGGCGGATCGATCAGTCAGGGGTG
AAATCCC
AGGGCTCAACCCTGGAACTGCCTTTGATACTGTCGATCTGGAGTATGGAAGAGGT
GAGTGGA
ATTCCGAGTGTAGAGGTGAAATTCGTAGATATTCGGAGGAACACCAGTGGCGAAG
GCGGCTC
ACTGGTCCATTACTGACGCTGAGGTGCGAAAGCGTGGGGAGCAAACAGGATTAGA
TACCCTG
GTAGTCCACGCCGTAAACGATGAATGTTAGCCGTCGGGCAGTAT

Isolate 52
GTCGAACGGCAGCACGGAGCTTGCTCTGGTGGCGAGTGGCGAACGGGTGAGTAAT
ATATCGG
AACGTACCCTGGAGTGGGGGATAACGTAGCGAAAGTTACGCTAATACCGCATACG
ATCTCAG
GATGAAAGCAGGGGACCTTCGGGCCTTGTGCTCCTGGAGCGGCCGATATCTGATT
AGCTAGT
TGGTGGGGTAAAGGCCTACCAAGGCATCGATCAGTAGCTGGTCTGAGAGGACGAC
CAGCCAC
ACTGGAACTGAGACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGGAATTTTG
GACAATG
TGCGCAAGCCTGATCCAGCAATGCCGCGTGAGTGAAGAAGGCCTTCGGGTTGTAA
AGCTCTT
TTGTCAGGGAAGAAAAGGCTGCGGCTAATATCTGCGGCTCATGACGGTACCTGAA
GAATAAG
CACCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGTGCAAGCGTTAAT
CGGAATT
ACTGGGCGTAAAGCGTGCGCAGGCGGTTTTGTAAGACTGTCGTGAAATCCCCGGG
CTCAACC
TGGGAATGGCGATGGTGACTGCGAGGCTAGAGTTTGGCAGAGGGGGGTAGAATTC
CACGTGT
AGCACTGAAATGCGTAGATATGTGGAGGAACACCGATAGCGAAGGCAGCCCCCTG
GGTCAAA
ACTGACGCTCATGCACGAAAGCGTGGGGAGCACACAGGATTAGATACCCTGGTAG
TCCACGC
CCTAAACGATGTCTACTAGTTGTCGGGTCTTAATTGACTCTGATAACGCAGCTCA
CGCGTGA AGT
Isolate 53
CGGCAGACGGGTGAGTAACGCGTGGGAATCTACCCTTGACTACGGAATAACGCAG
GGAAACT
TGTGCTAATACCGTATGTGTCCTTCGGGAGAAAGATTTATCGGTCAAGGATGAGC
CCGCGTT
GGATTAGCTAGTTGGTGGGGTAAAGGCCTACCAAGGCGACGATCCATAGCTGGTC
TGAGAGG
ATGATCAGCCACATTGGGACTGAGACACGGCCCAAACTCATACGGGAGGCAGCAG
TGGGGAA
TATTGGACAATGGGCGCAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCC
CTAGGGT
TGTAAAGCTCTTTCACCGGAGAAGATAATGACGGTATCCGGAGAAGAAGCCCCGG
CTAACTT
CGTGCCAGCAGCCGCGGTAATACGAAGGGGGCTAGCGTTGTTCGGAATTACTGGG
CGTAAAG

CGCACGTAGGCGGATCGATCAGTCAGGGGTGAAATCCCAGGGCTCAACCCTGGAA
CTGCCTT
TGATACTGTCGATCTGGAGTATGGAAGAGGTGAGTGGAATTCCGAGTGTAGAGGT
GAAATTC
GTAGATATTCGGAGGAACACCAGTGGCGAAGGCGGCTCACTGGTCCATTACTGAC
GCTGAGG
TGCGAAAGCGTGGGGAGCAAACAGGATTAGAT

Isolate 54
GGAGCGGCAGACGGGTGAGTAACGCGTGGGAATCTACCCTTGACTACGGAATAAC
GCAGGGA
AACTTGTGCTAATACCGTATGTGTCCTTCGGGAGAAAGATTTATCGGTCAAGGAT
GAGCCCG
CGTTGGATTAGCTAGTTGGTGGGGTAAAGGCCTACCAAGGCGACGATCCATAGCT
GGTCTGA
GAGGATGATCAGCCACATTGGGACTGAGACACGGCCCAAACTCCTACGGGAGGCA
GCAGTGG
GGAATATTGGACAATGGGCGCAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAA
GGCCCTA
GGGTTGTAAAGCTCTTTCACCGGAGAAGATAATGACGGTATCCGGAGAAGAAGCC
CCGGCTA
ACTTCGTGCCAGCAGCCGCGGTAATACGAAGGGGGCTAGCGTTGTTCGGAATTAC
TGGGCGT
AAAGCGCACGTAGGCGGATCGATCAGTCAGGGGTGAAATCCCAGGGCTCAACCCT
GGAACTG
CCTTTGATACTGTCGATCTGGAGTATGGAAGAGGTGAGTGGAATTCCGAGTGTAG
AGGTGAA
ATTCGTAGATATTCGGAGGAACACCAGTGGCGAAGGCGGCTCACTGGTCCATTAC
TGACGCT
GAGGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATA

Outgroup Bradyrhizobium sp. gene for 16S rRNA, partial sequence (Accession FJ544532.1)
TGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCCATGCCGCGTGAGTGAT
GAAGGCC
CTAGGGTTGTAAAGCTCTTTTGTGCGGGAAGATAATGACGGTACCGCAAGAATAA
GCCCCGG
CTAACTTCGTGCCAGCAGCCGCGGTAATACGAAGGGGGCTAGCGTTGCTCGGAAT
CACTGGG
CGTAAAGGGTGCGTAGGCGGGTCTTTAAGTCAGGGGTGAAATCCTGGAGCTCAAC
TCCAGAA
CTGCCTTTGATACTGAAGATCTTGAGTTCGGGAGAGGTGAGTGGAACTGCGAGTG
TAGAGGT
GAAATTCGTAGATATTCGCAAGAACACCAGTGGCGAAGGCGGCTCACTGGCCCGA
TACTGAC
GCTGAGGCACGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACG
CCGTAAA
CGATGAATGCCAGCCGTTAGTGGGTTTACTCACTAGTGGCGCAGCTAACGCTTTA
AGCATTC
CGCCTGGGGAGTACGGTCGCAAGATTA

Type strain *Rhizobium* leguminosarum bv. trifolii clone

(Accession Rt621KX486575.1)
GAGCGGCAGACGGGTGAGTAACGCGTGGGAATCTACCCTTGACTACGGAATAACG
CAGGGAA
ACTTGTGCTAATACCGTATGTGTCCTTCGGGAGAAAGATTTATCGGTCAAGGATG
AGCCCGC
GTTGGATTAGCTAGTTGGTGGGGTAAAGGCCTACCAAGGCGACGATCCATAGCTG
GTCTGAG
AGGATGATCAGCCACATTGGGACTGAGACACGGCCCAAACTCCTACGGGAGGCAG
CAGTGGG
GAATATTGGACAATGGGCGCAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAG
GCCCTAG
GGTTGTAAAGCTCTTTCACCGGAGAAGATAATGACGGTATCCGGAGAAGAAGCCC
CGGCTAA
CTTCGTGCCAGCAGCCGCGGTAATACGAAGGGGGCTAGCGTTGTTCGGAATTACT
GGGCGTA
AAGCGCACGTAGGCGGATCGATCAGTCAGGGGTGAAATCCCAGGGCTCAACCCTG
GAACTGC
CTTTGATACTGTCGATCTGGAGTATGGAAGAGGTGAGTGGAATTCCGAGTGTAGA
GGTGAAA
TTCGTAGATATTCGGAGGAACACCAGTGGCGAAGGCGGCTCACTGGTCCATTACT
GACGCTG
AGGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGT
AAACGAT
GAATGTTAGCCGTCGGGCAGTATACTGTTCGGTGGCGCAGCTAACGCATTA

A 1. 4 Bio assays A 1.4.1 Biofilm statistical output

GenStat Release 16.1 (PC/Windows 7) 11 November 2015 09:14:36
Copyright 2013, VSN International Ltd.
Registered to: Lincoln Uni

A 1.4.1.1 24 hour incubation time

LSD= 0.05
P<0.001

Mean		
47	0.2305	a
36	0.2439	ab
30	0.2530	abc
46	0.2567	abcd
33	0.2600	bcde
32	0.2612	bcde
44	0.2621	bcde
41	0.2646	bcde
48	0.2671	bcde
39	0.2682	bcdef
31	0.2692	bcdef
43	0.2726	bcdef
50	0.2735	cdef

49	0.2747	cdef
37	0.2776	cdefg
34	0.2781	cdefg
38	0.2783	cdefg
54	0.2783	cdefg
35	0.2831	defgh
51	0.2839	defgh
45	0.2870	efgh
52	0.2893	efgh
40	0.2967	fgh
53	0.3043	gh
42	0.3097	h

A 1.4.1.2 48 hour incubation time

LSD= 0.05
P<0.001
Mean

39	0.2636	a
45	0.2709	ab
46	0.2723	ab
48	0.2730	ab
40	0.2745	ab
36	0.2832	abc
38	0.2871	abc
43	0.2874	abc
44	0.2952	abc
50	0.3014	abcd
37	0.3041	abcd
34	0.3077	abcde
30	0.3093	abcde
51	0.3188	abcde
49	0.3258	abcdef
47	0.3276	abcdef
41	0.3344	abcdef
32	0.3428	bcdef
42	0.3440	bcdef
31	0.3604	cdef
52	0.3801	def
35	0.3832	ef
33	0.3983	f
54	0.5040	g
53	0.5046	g

A 1.4.1.3 72 hour incubation time

LSD= 0.05
P<0.001

	Mean	
30	0.2315	a
38	0.2384	ab
39	0.2475	abc
40	0.2478	abc
51	0.2516	abc
50	0.2582	abc
45	0.2709	abcd
46	0.2723	abcd
48	0.2730	abcd
43	0.2873	abcd
49	0.2882	abcd
44	0.2952	abcd
41	0.3025	abcd
47	0.3276	abcd
32	0.4071	abcde
37	0.4293	abcde
36	0.4581	abcde
35	0.4593	abcde
54	0.5176	bcde
34	0.5198	bcde
31	0.5255	cde
33	0.5461	de
52	0.6342	e
53	0.6424	e
42	1.0645	f

A 1.4.1 PEG stats

A 1.4.1.1 4hour incubation time 60% PEG

LSD =0.05
P<0.001

Mean		
31	0.00800	a
34	0.01200	ab
32	0.01300	abc
35	0.01350	abc
33	0.01750	abc
37	0.01800	abc

36 0.02150
abcd 49
0.03050 bcde
38 0.03200
cdef 52
0.04100 defg
41 0.04450 efg
47 0.04450 efg
53 0.04450 efg
43 0.04500 efg
51 0.04500 efg
54 0.04600 efg
39 0.04750 efgh
42 0.04800 efgh
44 0.04900 efgh
45 0.05050 fgh
50 0.05250 gh
48 0.05450 gh
46 0.05700 gh
40 0.06700 h

A 1.4.1.2 6hour incubation time 60% PEG

LSD =0.05
P=0.003

Mean
36 0.00650 a
37 0.00850 a
44 0.00900 a
32 0.01250 ab
52 0.01250 ab
54 0.01650 abc
33 0.01700 abc
38 0.01750 abc
48 0.01750 abc
53 0.01750 abc
51 0.01900 abc
50 0.01950 abc
45 0.02050 abc
49 0.02050 abc
47 0.02250 abc
46 0.02400 abc
42 0.02600 abc
35 0.02600 abc
41 0.02650 abc
31 0.03150 bc
39 0.03200 bc

43 0.03650 c
40 0.03750 c
34 0.07550 d

A 1.4.1.3 14hour incubation time 60% PEG

LSD =0.05
P<0.001
Mean
33 0.00650 a
52 0.00800 ab
54 0.00850 abc
45 0.01300 abcd
32 0.01400 abcde
53 0.01400 abcde
31 0.01500 bcdef
34 0.01650 cdef
39 0.01700 def
40 0.01700 def
37 0.02000 def
49 0.02000 def
35 0.02050 def
51 0.02050 def
50
0.02150 ef 38

0.02300 fg
36 0.03050 gh
48 0.03250 h
42 0.03350 h
47 0.03400 h
46 0.03600 h
41 0.03800 h
43 0.03850 h
44 0.05800 i

A 1.4.1.4 Combined incubation time 60%

LSD =0.05
P<0.001
Mean
54 2.012 a
33 2.028 a

36	2.085	ab
52	2.094	ab
35	2.112	ab
31	2.116	ab
37	2.118	ab
32	2.143	abc
38	2.155	abc
47	2.161	abc
50	2.192	abc
45	2.196	abc
51	2.207	abcd
53	2.213	abcd
44	2.237	abcd
48	2.237	abcd
34	2.270	bcde
43	2.274	bcde
46	2.277	bcde
49	2.291	bcde
40	2.313	bcde
39	2.368	cde
41	2.440	de
42	2.486	e

A 1.4.1.4 6 hour incubation time 50% PEG

LSD= 0.05
P<0.001
Mean

45	0.00750	a			
51	0.00800	a			
49	0.01350	ab			
47	0.01450	ab			
50	0.01550	abc			
48	0.01950	abcd			
53	0.02050	abcde			
46	0.03350	abcdef			
54	0.03650	abcdefg			
32	0.04650	bcdefgh			
52	0.04950	cdefghi			
31	0.05200	defghi			
37	0.05250	defghi			
38	0.05450	efghi			
36	0.05550	fghi			
44	0.05650	fghi			
35	0.06200	fghij			
42	0.06350	fghij			
43	0.06500	fghij	40	0.07000	ghij
34	0.07350	hij			

33	0.07400	hij
41	0.08150	ij
39	0.09550	j

A 1.4.1.5 14 hour incubation time 50% PEG

LSD= 0.05
P<0.001

Mean		
35	0.00450	a
53	0.00450	a
44	0.00550	ab
40	0.00750	abc
42	0.00750	abc
38	0.00900	abcd
41	0.01000	abcde
48	0.01000	abcde
43	0.01050	abcde
31	0.01250	abcdef
54	0.01250	abcdef
50	0.01300	abcdef
51	0.01300	abcdef
45	0.01500	bcdef
47	0.01500	bcdef
49	0.01550	cdef
46	0.01700	cdef
36	0.01850	def
39	0.01950	ef
52	0.01950	ef
37	0.02200	f
34	0.03200	g
33	0.03300	g
32	0.03900	g

A 1.4.1.6 48 hour incubation time 50% PEG

LSD =0.05
P<0.001

Mean		
46	0.00400	a
37	0.00550	ab
36	0.00600	ab
49	0.00650	ab
44	0.01050	abc
45	0.01150	abc
40	0.01200	abcd
41	0.01200	abcd
53	0.01550	abcde
35	0.01550	abcde

47	0.01600	abcde
39	0.01650	abcdef
50	0.01750	abcdef
42	0.01800	bcdef
52	0.01800	bcdef
54	0.01800	bcdef
51	0.01900	bcdefg
48	0.02100	cdefg
33	0.02550	defg
32	0.02900	efgh
34		
0.03000	fgh	43
0.03250	ghi	
31	0.04000	hi
38	0.04500	i

A 1.4.1.7 combine incubation time 50%

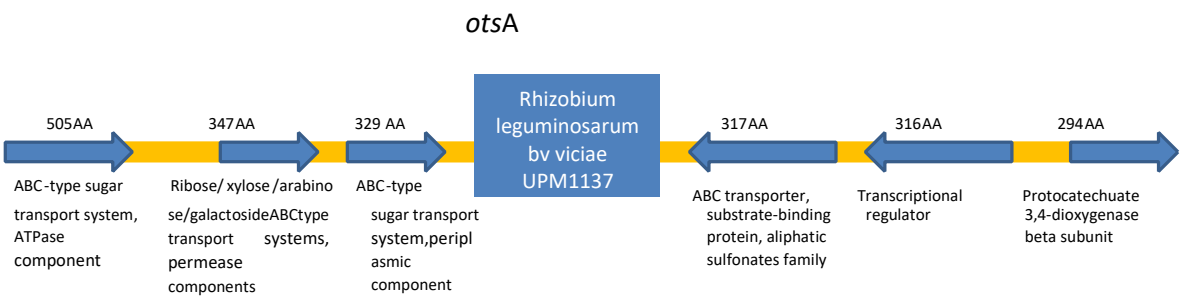
Mean		
49	1.105	a
47	1.142	ab
53	1.220	abc
45	1.227	abc
51	1.231	abcd
50	1.238	abcd
48	1.241	abcd
35	1.271	abcde
36	1.295	bcde
46	1.313	bcdef
54	1.321	cdef
44	1.350	cdefg
37	1.356	cdefgh
52	1.377	cdefgh
40	1.380	cdefgh
42	1.384	cdefgh
41	1.403	defghi
43	1.419	efghi
38	1.485	fghij
31	1.509	ghij
39	1.515	ghij
33	1.531	hij
32	1.560	ij
34	1.599	j

A 2 Chapter 3

Table 5.1 Primer design

IUPAC NUCLEOTIDE CODE	BASE
A	Adenine
C	Cytosine
G	Guanine
T (OR U)	Thymine (or Uracil)
R	A or G
Y	C or T
S	G or C
W	A or T
K	G or T
M	A or C
B	C or G or T
D	A or G or T
H	A or C or T
V	A or C or G
N	any base
. OR -	gap

A 2.2 Contigs



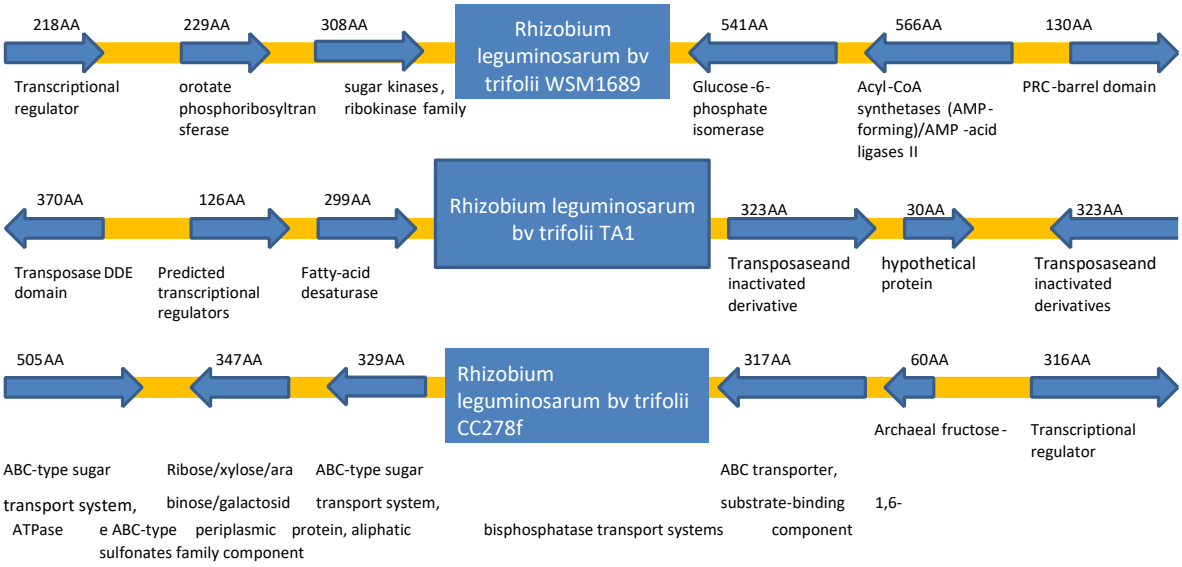


Figure 5.1 This figure shows the α-α-Trehalose-6-phophate synthase (otsA) gene contiguous sequences for four strains of R. leguminosarum. The orientations of the genes are illustrated by the arrow direction, right facing arrows are on the 5'- 3' sense strand and left pointing arrows are located on the antisense strand. The relative peptide strand is also illustrated

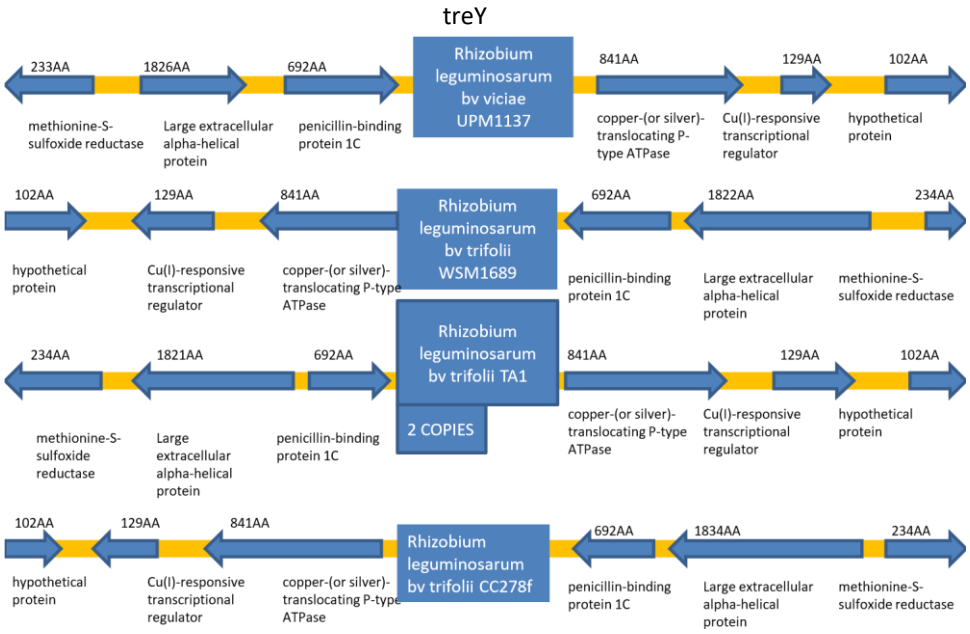


Figure 5.2 This figure shows the maltooligosyl trehalose synthase (treY) gene contiguous sequences for four strains of *R. leguminosarum*. The orientations of the genes are illustrated by the arrow direction, right facing arrows are on the 5'- 3' sense strand and left pointing arrows are located on the antisense strand. The relative peptide strand is also illustrated

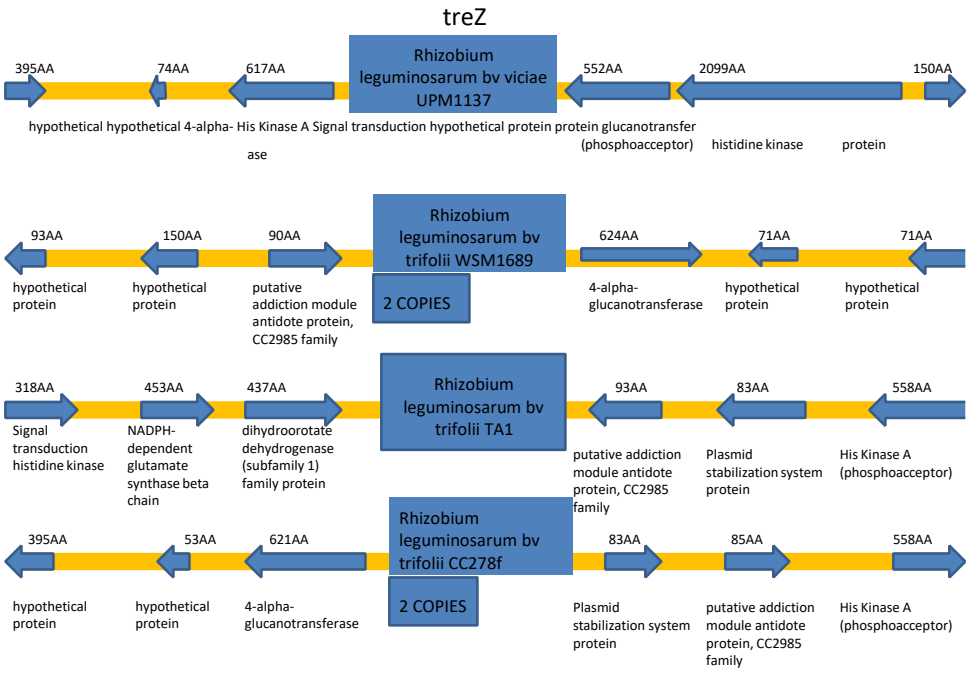


Figure 5.3 This figure shows the maltoologosyltrehalose trehalohydrolase (treZ) gene contiguous sequences for four strains of *R. leguminosarum*. The orientations of the genes are illustrated by the arrow direction, right facing arrows are on the 5'- 3' sense strand and left pointing arrows are located on the antisense strand. The relative peptide strand is also illustrated.

A2.3 Reference amino acid sequences

A **2.2.3.1 otsB** reference genome amino acid sequence

Rhizobium leguminosarum bv *trifolii* CC278f
MQDDDATQEMLSLVLEEPDHWAMFLDIDGTLLNLAPTPDAIEVPEALPGQ
LHRLSNKLGGALALVTGRSLAYADELFKPFEFPTAGLHGAEIRNAAGMQT
VEATAEFQALKHALTAEAEHYPGVLIEDKGAAVAAHYRLAPEYEKVLEDR
MHHYAELAGPNWALQLGKMVFELRPARSSKGDALERFFQSDPFKNRCPIT
IGDDLTDESMFAIANARGGVSVRVGAIGAPSCATSRLSSSALVRNVIAAL
AA

Rhizobium leguminosarum bv *trifolii* TA1
MQDDDATQEMLSLVLEEPDHWAMFLDIDGTLLNLAPTPDAIEVPEALPGQ
LHRLSNKLGGALALVTGRSLSYADELFKPFEFPTAGLHGAEIRNAAGMQT
VEATAEFQALKHALTAEAEHYPGVLIEDKGAAVAAHYRLAPEYEKVLEDR
MHHYAELAGPNWALQLGKMVFELRPARSSKGDALERFFQSDPFKNRCPIT
IGDDLTDESMFAIANARGGVSVRVGAIGAPSCATSRLSSSALVRNVIAAL
AA

Rhizobium leguminosarum bv viciae UPM1137
VQDDDATQEMLSLVLEEPDHWAMFLDIDGTLLNLAPTPDAIEVPEALPGQ
LHRLSNKLGGALALVTGRSLAYADALFKPFAFPTAGLHGAEIRNAAGMQT
VEATAEFQALKHALTAEAEHYPGVLIEDKGAAVAAHYRLAPEYEKVLEDR
MHHYAELAGPNWALQLGKMVFELRPARSSKGDALERFFQSDPFKNRCPIT
IGDDLTDESMFAIANARGGVSVRVGAIGAPSCATSRLSSSALVRNVIAAL
AA

Rhizobium leguminosarum bv *trifolii* WSM1689
MVLEEPDRWAMFLDIDGTLLNLAPTPDAIEVPEALPGQLHRLSNKLGGAL
ALVTGRSLAYADALFKPFAFPTAGLHGAEIRNAAGMQTIEATPEFQALKH
ALTAEAEHYPGVLIEDKGAAVAAHYRLAPEYEKVLEERMHHYAEIAGPNW
ALQLGKMVFELRPARSSKGDALERFFQSDPFKNRSPITIGDDLTDESMFA
IANARGGVSVRVGAIGAPSCATSRLSSAALVRNVIAALAA

A 2.2.3.2 *ots*A reference genome amino acid sequence

Rhizobium leguminosarum bv *trifolii* CC278f

MARLVVVSNRVPIPDKDGVAPAGGLAVALQAALQERGGIWMGWSGKSSGSREPER
LSRQQQG
NITYAVTDLTDTDVEEYYHGFANRVLWPICHYRLDLAEYGRKEMAGYFRVNRFFA
QRLAPLI
EPDDVIWVHDYHLIPLAAELRRIGLKNRIGFFLHIPWPSADVLVTMPVHEEIMRG
LSHYDLL
GFQTDHDLQNFAGYLTREGIGNDLGDGLFDSHGRTFKSAAYPIGIETAAFAELAG
KAWSNIL
VRKARQSIEGRDLIIGVDRLDYSKGITQRLDAFERFITTNPAYQNKVTYLQITPK
SRSEVPE
YEQMQRMVAEQAGRVNGAIGTVDWVPIRYINRSSSRNVLAGLYRLATIGFVTPLR
DGMNLVA
KEYVAAQDPDRPGVLVLSRFAGAARELKGALLVNPYDVEGTANALARALTMSIEE
RRERWRG
MMDHLLEYDVSRWCNDFLRDLTAKTAAAN

Rhizobium leguminosarum bv *trifolii* TA1
MSRLIVVSNRVPMPAKDGSAAAGGLAVALQAALQERGGIWMGWSGESSGDREPGP
LSQLQKG
NITYALTDLTDTDVEEYYRGFANRVLWPICHYRLDLAEYGRKEMAGYFRVNRFFA
HRLAPMI
EPDDIIWVHDYHLIPLAAELRQMGLKNRIGFFLHIPWPPADILVTMPVHEEIMRG
LSHYDLV
GFQTDYDLQNFAGYLRREGIGDDLGNGLFDSHGRIFKAGAYPIGIETAAFAEFAE
RAANNIM
VQKTRRSVEGRDMIIGVDRLDYSKGIIQRLEAFERFITNNPAYQNKVTYLQITPK
SRSEVPE
YEHMQNMVAEQAGRVNGAIGTVDWVPIRYVNRSISRNVLAGLYRLATIGLVTPLR
DGMNLVA
KEYVAAQDPDRPGVLVLSRFAGAARELKGALLVNPYDVEGTANAIAKGLAMSLEE
RRDRWSM
MMEHLLSHDVSLWCENFLRDLVLAPELRPERDPG

Rhizobium leguminosarum bv *viciae* UPM1137
MARLVVVSNRVPIPDKDGVAPAGGLAVALQAALQERGGIWMGWSGKSSGSREPER
LSRRQQG
NITYAVTDLTDTDVEEYYHGFANRVLWPVCHYRLDLAEYGRKEMAGYFRVNRFFA
QRLAPLI
EPDDVIWVHDYHLIPLAAELRRIGLKNRIGFFLHIPWPSADVLVTMPVHEEIMRG
LSHYDLL
GFQTDHDLQNFAGYLTREGIGNDLGDGLFDSHGRTFKAAAYPIGIETAAFAELAG
KASANIM
VRKARQSVEGRDLIIGVDRLDYSKGITQRLDAFERFITNNPAYQNKVTYLQITPK
SRSEVPE
YEQMQRMVAEQAGRVNGAIGTVDWVPIRYINRSSSRNVLAGLYRLATIGFVTPLR
DGMNLVA

KEYVAAQDPDRPGVLVLSRFAGAARELKGALLVNPYDVEGTANALAKALTMSIEE
RRERWRG
MMDHLLEHDVSRWCNDFLRDLIVKTPAAN

Rhizobium leguminosarum bv *trifolii* WSM1689
MSRLIVVSNRVPMPAKDGSAAAGGLAVALQAALQERGGIWMGWSGESSGDREPGP
LSQLQKG
NITYALTDLTDTDVEEYYRGFANRVLWPICHYRLDLAEYGRKEMAGYFRVNRFLA
HRLAPMI
EPDDIIWVHDYHLIPLAAELRQMGLQNRIGFFLHIPWPPADILVTMPVHEEIMRG
LSHYDLV
GFQTDYDLQNFAGYLRREGIGDDLGNGLFDSHGRIFKAGAYPIGIETAGFAEFAE
RAANNIM
VQKTRRSIEGRDMIIGVDRLDYSKGIIQRLEAFERFLTSNPGYQNKVTYLQVTPK
SRSEVPE
YEHMQKMVAEQAGRVNGAIGTVDWVPIRYVNRSISRNVLAGLYRLATIGLVTPLR
DGMNLVA
KEYVAAQDPDRPGVLVLSRFAGAARELKGALLVNPYDVEGTANAIARGLAMSLEE
RRDRWSM
MMEHLLSHDVSLWCKDFLRDLVPAPELRSQRDPG

A 2.2.3.3 *tre*Y reference genome amino acid sequence

Rhizobium leguminosarum bv *trifolii* CC278f
MTLPTATYRIQFRNGMTFDGACDLVPYLKTLGISHLYASPIFTAVSGSTHGYDVT
DANEIDP
ALGGRAGFERLTESLASSGMGLILDIVPNHMAASSENGWWRDVLAFGRQSAYFSH
FDIDWSE
PLTLPQLGQDFEGALAGGELRITLDETHGNFAFGYFETLLPLNPTSYGAIANRLD
DPVATRM
AEAAAVTSGENFNRAMRDILFEGGDRAVLRQKLDDVSADRDFMRSLHEGQHWRLT
HWKAAAR
HLSYRRFFEVTGLVGTRVESPAVFEDMHRLVIELVRHGKVQGLRIDHVDGLAEPR
AYLDRLR
EAAGPDTYIVVEKIIGASEVLPVSWPVAGTTGYEFIAALSELFIDGGGLRILDDA
YRRLAAE
TADLEESRRRAKRLMVERNFAGETGRLVSIAADIFPEVKRDEIASALSELLITFP
VYRTYGD
GGPLSWQDSAVLAATASQAMARLDDRRACDHVLKLLEGKVEGDAAHEFRIRFQQL
SGPVMAK
ATEDTLFYRYNRLLAANEVGGEPGKAPGGPEEFHRRMAERARLQPHGLSASATHD
TKRGEDA
RARLYALSEGADVFAQAVERWRDMNRPWLKDLPEGAAPEPNVEWMLYQALAGIWP
EDFDRGR
TEELRERFTDYAVKAVREAKLRSDWTEQDADYEEAVTTYAAALVSPDNDVFLEDF
ERVLQPF
IAAGYLNSLSQTLLKLTAPGIPDIYQGAESFDFSLVDPDNRRPVDHEQLKVWLDE
AGPIAKL
QAAALKQRIVGIGLQLRQRHQDLFARGDYLPLKVTGNRRDHVLAFARVQKSDFAI
IAAPRLM
FGWLDPGVLFAGPEFWEDTAIAVPSPLHGLKADLLTGKTVEPGGSISVAALLGSQ
PVGLITP

I

Rhizobium leguminosarum bv *trifolii* TA1
MTLPTATYRIQFRNGMTFDRARGLVPYLKTLGISHLYASPIFTAVSGSTHGYDVT
DANEIDP
VLGGRASFDRLTESLASAGMGLILDIVPNHMAASPENGWWRDVLTFGRQSAYFRH
FDIDWSE
PLTLPQLGQDFDRALADGELRIALDETLGNFVFRYFGTLLPLKPGSYGAIANRLD
DPVAMRM
AEAAAVTSGDNFNRAMRDILFEGGDRAVLEQKLDDVSADRDFMRSLHEAQHWRLT
HWKAAAR
HLSYRRFFEVTGLVGTRVEDPAVFEDMHRLVIELVRHGKVQGLRIDHVDGLAEPK
AYLDRLR
EAAGPDTYIVVEKILGAGEVLPESWPIAGTTGYEFIAALSELFIDGGGLRILDDA
YRSIAAD
TGDLEEGRRIAKRLMVERNFAGETDRLVSIAAGIFPELKRDEIATALSELLITFP
VYRTYGD
GGPLSWQDSAVLAVTASQAMARLDDRRACDHVLKLLEGKVEGDAAHEFRIRFQQL
SGPVMAK
ATEDTLFYRYNRLLAANEVGGEPGKAPGGPEEFHRRMAERARLQPHGLSASATHD
TKRGEDA
RARLYALSEGADVFAQAVERWREMNRPWLKDLPEGAAPEPNVEWMLYQALAGIWP
EDFDRGQ
TEELRERFTDYALKAVREAKLRSDWTEQDTAYEEAVTTYAAALVSPDNDVFLEDF
ERVLQPF
IAAGYLNSLSQTLLKLTAPGIPDIYQGAEGFDFSLVDPDNRRPVDHERLTAWLDE
AGPIAKL
QAAALKQRLVGIGLQLRQRQAELFARGDYLPLKVTGARRDHVLAFARVHKDDSAI
IAAPRLM
FGWLDPGVLFAGPEFWEDTAIAVPSPLHGLKADLLTGKTVEPGGSISVATLLGSQ
PVGLITP
TEGQPTQRSYP

Rhizobium leguminosarum bv *trifolii* TA1
MTLPTATYRIQFRNGMTFDRARGLVPYLKTLGISHLYASPIFTAVSGSTHGYDVT
DANEIDP
VLGGRASFDRLTESLASAGMGLILDIVPNHMAASPENGWWRDVLTFGRQSAYFRH
FDIDWSE
PLTLPQLGQDFDRALADGELRIALDETLGNFVFRYFGTLLPLKPGSYGAIANRLD
DPVAMRM
AEAAAVTSGDNFNRAMRDILFEGGDRAVLEQKLDDVSADRDFMRSLHEAQHWRLT
HWKAAAR
HLSYRRFFEVTGLVGTRVEDPAVFEDMHRLVIELVRHGKVQGLRIDHVDGLAEPK
AYLDRLR
EAAGPDTYIVVEKILGAGEVLPESWPIAGTTGYEFIAALSELFIDGGGLRILDDA
YRSIAAD
TGDLEEGRRIAKRLMVERNFAGETDRLVSIAAGIFPELKRDEIATALSELLITFP
VYRTYGD
GGPLSWQDSAVLAVTASQAMARLDDRRACDHVLKLLEGKVEGDAAHEFRIRFQQL
SGPVMAK
ATEDTLFYRYNRLLAANEVGGEPGKAPGGPEEFHRRMAERARLQPHGLSASATHD
TKRGEDA

RARLYALSEGADVFAQAVERWREMNRPWLKDLPEGAAPEPNVEWMLYQALAGIWP
EDFDRGQ
TEELRERFTDYALKAVREAKLRSDWTEQDTAYEEAVTTYAAALVSPDNDVFLEDF
ERVLQPF
IAAGYLNSLSQTLLKLTAPGIPDIYQGAEGFDFSLVDPDNRRPVDHERLTAWLDE
AGPIAKL
QAAALKQRLVGIGLQLRQRQAELFARGDYLPLKVTGARRDHVLAFARVHKDDSAI
IAAPRLM
FGWLDPGVLFAGPEFWEDTAIAVPSPLHGLKADLLTGKTVEPGGSISVATLLGSQ
PVGLITP
TEGQPTQRSYP

Rhizobium leguminosarum bv *viciae* UPM1137
MTLPTATYRIQFRNGMTFDRACDLVPYLKMLGISHLYASPIFTAVSGSTHGYDVT
DANEIDP
VLGGRAGFERLTESLASAGMGLILDIVPNHMAASPENGWWRDVLISGRQSAYFSH
FDIDWSE
PLTLPQLGQDFEGALAGGELRIALDETHGNFAFGYFETLLPLNAASYGAIANRLD
DPVATRM
AEAAAVTSGENFNRVMRDILFEGGDRAALQQKLNDVSADRDFMRSLHEGQHWRLT
HWKGAAR
HLSYRRFFEVTGLVGTRVEDPPVFEDMHRLVIELVRHGRVQGLRIDHVDGLAEPK
AYLDRLR
EAVGPDTYIVVEKILGASEVLPESWPVAGTTGYEFIAALSELIDAGGLRILDDAY
RSIAVET
GDLEAGRRIAKELMVERNFAGETSRLVSIAAGIFPELKRDEIATALSELLITFPV
YRTYGDG
GPLSWQDSAVLAATASQAMARLDQRALDQVLKLLEGKVEGDAAHEFRIRFQQLSG
PVMAKAT
EDTLFYRYNRLLAANEVGGEPGKAPGGPEEFHRRMAERARLQPHGLSTSATHDTK
RGEDARA
RLYALSEGADVFAQAVERWRDMNRPWLKDLPDGAAPEPNVEWMLYQALAGIWPED
FDRGRTE
ELRERLTDYAVKAVREAKLRSDWTEQDAAYEEAVKTYAAALVSPDNDVFLEDFER
VLQPFIA
AGYLNSLSQTLLKLTAPGIPDIYQGAESFDFSLVDPDNRQPVDHERLKAWLDEAG
PIAKLQA
AALKQRIVGIGLQLRQRHADLFARGDYLPLKVTGNRRDHVLAFARVQKSDFAIIA
APRLMFG
WLDPGVLFAGPEFWEDTAIAVPSPLHGLKADLLTAKTIEPGGSISVAALLGSQPV
GLITPI

Rhizobium leguminosarum bv *trifolii* WSM1689
MTLPTATYRIQFRNGMTFDRACDLVPYLKTLGISHLYASPIFTAVSGSTHGYDVT
DANDIDP
ALGGRAGFERLTESLASAGMGLILDIVPNHMAASPENGWWRDVLTFGRQSAYFSH
FDIDWSE
PLTLPQLGQDFEGALAGGELRITLDETHGNFAFGYFETLLPLNPTSYGAIAHRLD
DPVATRM
AEAGAVTSGENFNRAMRDILFEGGDRAVLRQKLEEVSSDGDFVRSLHEAQHWRLT
HWKEAAR
HLSYRRFFEVTGLVGIRVEDPPVFEDMHRLVIELVRHGKVQGLRIDHVDGLAGPT
AYLDRLR

EAAGPDTYIVVEKILGTGEVLPESWPVAGTTGYEFIAALSELFIDGGGLRILDDA
YRSIAGE
TGNLEEGRRIAKRLMVERNFAGETDRLVSIAAGIFPEVKRDEITTALSELLIAFP
VYRTYGD
GGPLSWQDSAVLAATASQAMARLDDRRACDHVLTLLEGKVEGDAAHDFRIRFQQL
SGPVMAK
ATEDTLFYRYNRLLAANEVGGEPGKTPGGPEEFHRRMAERARLQPHGLSASATHD
TKRGEDA
RARLYALSEGADVFAQAVERWREMNRPWLKDLPDGAAPEPNVEWMLYQALAGIWP
EDFDRGQ
TEELRARFADYAVKAVREAKLRSNWTEQNQAYEEAVTTYAAALVSPDNDVFLEDF
ERVLQPF
IAAGYLNSLSQTLLKLTAPGIPDIYQGAEGFDFSLVDPDNRRLVDHQRLTAWLAE
PGPIAKL
QAAALKQRLVGIGLQLRQRHVDLFAGGDYLPLKVTGNRRDHALAFARVHKGDFAI
IAAPRLM
FGWLDPGVLFAGPEFWEDTAIAIPSPLHGLKADLVTGKTIEPGGSISVAALLGSQ
PVGLITP S

A 2.2.3.4 *treZ* reference genome amino acid sequence

Rhizobium leguminosarum bv *trifolii* WSM1689
MHHHSSSEKSKPMPPRSRRRLPIGAEVDAGGVSFRLWAPARERCFLAIEGNSEHE
MTAEDGG
YFCLHLPGLDAGTRYQFHFGDADDLLADPASRFQPDGPSGPSVVVDPARYQWSDH
DWQGMPA
IGAALYEMHIGTFTREGTFAAAREKLERLRAIGINCLEVMPINEFEGEFGWGYDG
TLLYAPT
RLYGTPDDVRGFVDEAHRIGIGVILDVVYNHFGKGERFCEFTPDYFTERYSNEWG
KSINFDG
SNSRGVREYVAKNAAYWIDEFHFDGLRIDATQALFDSSHEHIITVIAREARAAAG
QRQIYLV
SENEPQQTNMVRPADVGGHGLDALWNDDFHRSATVALTGRSEAYYHDHRGTAQEL
VSAAKYG
CLFQGQRYDWQDKPRGTAGLDLKPWNFVNFLQNHDQVANSGTGARIGQITSPARL
RALTALQ
LLGPQTPMLFQGQEFGSSSPFYYFADHQGEIADIVRQGRRSFISQFPNLRDEELI
RQLADPC
ARTTFEKVKLDWAEWERNAALVLLHRDLLNLRQTNKAFSRQACARDGQMDGNILS
SSAFLLR
FFAGKPSDERILLINFGNDLVIDSLPDPLFAPPKAHQWHLSWSSEDAAYGGSGRR
PCDFRKR
WVLNGDIALVLAPLKLQNRQNASAMPIEDW
QAGILRAGDPGT

Rhizobium leguminosarum bv *trifolii* WSM1689

MTFGPAFTEKGILFRLWAPLHESVSLNLEGADPRPMQAVGNGWHQYTVADASVGT
RYRFVLP
DGLEIPDPASRFQPQDVHGPSEVVDLSTYRWKTSEWMGRPWEEMVIYEMHIGCFT
LEGTFKA

AIGRLDHLQQLGVTALQIMPLSEFPGRYSWGYDGVLPYAPDSSYGRPEDFMALVD
AAHQRGI
SIFLDVVYNHFGPDGNYIPSYAPLFTDHHKTPWGNGINYDGDGSEMIREFIIENA
IYWITEF
RLDGFRFDAVHAIKDDSAEHLLHALARRVRAAAGSRHVHLIVENEENDSDLLKRD
ENGEVRL
FTAQWNDDVHHVLHITATGETFGYYADYAGDAGKLGRALAEGFVFQGEHMPYRGA
SRGRPSG
HLPPTAFISFIQNHDQIGNRALGDRVLVSSPADVVKAVAAIYLLAPEIPMLFMGE
EWGAKEP
FPFFCDFDEDLNEKVRKGRREELSRLPGFDADDLLDPTAPSTFAAAKLDWSKLAS
SEALDFY
RMLLDLRHRRIVPLLKGAGAGNAVYRSAGSALAVDWTLAQNRRLHLRANLGAEAA
PLVSQQD
DGETIFRLGGSDGGNLAPWTVIWSIGEA

Rhizobium leguminosarum bv *viciae* UPM1137
MTFGPAFTEGGILFRLWAPLHESVSLKIEGADPWPMQATEDGWHRCTVAQAHTGT
LYRFVLP
DGSEVPDPASRFQPQDVRGPSEVVDLSTYRWTTSDWTGRPWEEMVIYEMHIGCFT
PEGTFKA
AIERLDHLQALGVTALQIMPLSEFPGRYSWGYDGVLPYAPDSSYGRPEDFMALVD
AAHLRGI
SVFLDVVYNHFGPDGNFIPAYAPLFTEHHKTPWGNGINYDGDGSEMIREFIIQNA
IYWITEF
RLDGFRFDAVHAIKDNSAEHLLHALARRVKAAAGDRLVHLIVENEENDSDLLKRD
ENGEVRL
FTAQWNDDVHHVLHITATGETFGYYADYVGDAGKLGRALAEGFVFQGEHMPYRGG
SRGRPSG
HLPPTAFISFIQNHDQIGNRALGDRVLASSPADVVKAVAAIYLLAPEIPMLFMGE
EWGAREP
FPFFCDFDEDLNEKVRKGRREELSRLPGFDADDLLDPTAPSTFAAAKLDWSKLAS
SDVFDFY
RTLLDLRHRRIVPLLKGAGAGNAVYRSAGSALAVDWTLAQNRRLHLRANLGTEAA
TLASQQD
DGETIFRLGGSDGGNLAPWTVIWKISEA

Rhizobium leguminosarum bv *trifolii* TA1
MTFGPAFTEEGILFRLWAPLHESVSLKIEGADPRPMQAVGNGWHHYTVADASVGT
RYRFVLP
DGLEIPDPASRFQPQDVHGPSEVVDLSFYRWKTSDWTGRPWEEMVIYEMHIGCFT
PEGTFKA
AIERLDHLQALGVTALQIMPLSEFPGRYSWGYDGVLPYAPDSSYGRPEDFMALVD
AAHQRGI
SVFLDVVYNHFGPDGNYIPAYAPLFTDHHKTPWGNGINYDGDGSEMIREFIIENA
IYWITEF
RLDGFRFDAVHAIKDDSAEHLLHALARRVRAAAGDWHVHLIVENEENDSDLLQRD
ENGEVKL
FTAQWNDDVHHVLHITATGETFGYYADYAGDAGKLGRALAEGFVFQGEHMPYRGG
SRGRPSG
HLPPTAFISFIQNHDQIGNRALGDRVLASSPADVVKAVAAIYLLAPEIPMLFMGE
EWGAKEP

FPFFCDFDEDLNEKVRKGRRQELSRLPGFDADDLLDPTAPSTFAAAKLDWSKLAS
SEVLGFY
RTLVDLRHRRIVPLLTGAGAGNAVYRSAGSALAVDWTLAQNRRLHLRANLGAEAA
PPVSQQG
DGETIFRLGGNDGDDLAPWTVIWSIGEA

Rhizobium leguminosarum bv *trifolii* CC278f
MLPQKLPKSHPNFIEGFPLTKMTFGPAFTEEGILFRLWAPLHESVSLKLEGAPPR
PMQAAED
GWHRCTVAQAHAGTLYRFVLPDGSEVPDPASRFQPQDVHGPSEVVDLSTYRWATS
DWTGRPW
EEMVIYEMHIGCFTPEGTFKAAIERLDHLQQLGVTALQIMPLSEFPGRYSWGYDG
VLPYAPD
SSYGRPEDFMALVDAAHRRGISVFLDVVYNHFGPDGNFIPAYAPLFTEHHKTPWG
NGINYDG
DGSEMIREFIIQNAIYWVTEFKLDGFRFDAVHAIKDDSTEHLLHALAHRIRAAAG
SRHVHLI
VENEENDSDLLKREENGEVTLFTAQWNDDVHHVLHITATGETFGYYADYAGDAGK
LGRALAE
GFVFQGEHMPYRGGSRGRPSVHLPPTAFISFIQNHDQIGNRALGDRVLASSPADV
VKAVAAI
YLLAPEIPMLFMGEEWGAREPFPYFCDFDEELNEKVRKGRREELSRLPGFDADDL
LDPTAPS
TFTAAKLDWSKLASSEVLGFYRTLLDLRHRRIVPLLKGAGAGNAVYRSAGSALAV
DWTLAQN
RRLHLRANLGTEAATLVSQQDDGETIFRLGGSDGRNLAPWTVIWSISEA

Rhizobium leguminosarum bv *trifolii* CC278f

MLPQKLPKSHPNFIEGFPLTKMTFGPAFTEEGILFRLWAPLHESVSLKLE
GAPPRPMQAAEDGWHRCTVAQAHAGTLYRFVLPDGSEVPDPASRFQPQDVHGPSE
VVDLSTY
RWATSDWTGRPWEEMVIYEMHIGCFTPEGTFKAAIERLDHLQQLGVTALQIMPLS
EFPGRYS
WGYDGVLPYAPDSSYGRPEDFMALVDAAHRRGISVFLDVVYNHFGPDGNFIPAYA
PLFTEHH
KTPWGNGINYDGDGSEMIREFIIQNAIYWVTEFKLDGFRFDAVHAIKDDSTEHLL
HALAHRI
RAAAGSRHVHLIVENEENDSDLLKREENGEVTLFTAQWNDDVHHVLHITATGETF
GYYADYA
GDAGKLGRALAEGFVFQGEHMPYRGGSRGRPSVHLPPTAFISFIQNHDQIGNRAL
GDRVLAS
SPADVVKAVAAIYLLAPEIPMLFMGEEWGAREPFPYFCDFDEELNEKVRKGRREE
LSRLPGF
DADDLLDPTAPSTFTAAKLDWSKLASSEVLGFYRTLLDLRHRRIVPLLKGAGAGN
AVYRSAG
SALAVDWTLAQNRRLHLRANLGTEAATLVSQQDDGETIFRLGGSDGRNLAPWTVI
WSISEA

A 2.2.3.5 *tre*S reference genome amino acid sequence

Rhizobium leguminosarum bv *trifolii* WSM1689

MDTLNPDGMNQPLWYKDAIIYQLHIKSFYDANGDGVGDFAGLHQKLDHIAALGVN
AIWLLPF
FPSPRRDDGYDIADYGSVSSDYGTVEDFQAFVDAAHQRNIRVIIELVINHTSDQH
PWFQRAR
QAPAGSPERDFYVWSDTDQKFPETRIIFIDTEKSNWTWDAVAGAYYWHRFYSHQP
DLNFDSP
LVMEELLRVMRFWLETGIDGFRLDAIPYLVEREGTINENLPETHAILKRIRAALD
ATHPGVM
LLAEANQWPEDTREYFGEGDECHMAFHFPLMPRMYMAIAKEDRFPITDILRQTPE
IPDNCQW
AIFLRNHDELTLEMVTDAERDYLWETYASDKRARINLGIRRRLAPLMERDRRRIE
LMNALLL
SMPGTPVIYYGDEIGMGDNIYLGDRDGVRTPMQWSPDRNGGFSRVDPARLVLPPV
ADPLYGF
EAVNVEAQSTDAHSLLNWTRKMLALRGRHPAFGRGSLRFLSPENRKILAYLREYE
GETLMCV
ANLSRLPQAVELDLSAFEGRVPIELTGMSPFPPIGQLTYLLTLPPYGFFWFQLEA
DADPPAW
RTAPPEQLPDLVTMVIRRGLLDLVDEPKLARTLSTEILPAYLAKRRWFGAKDQAL
QAARLIS
ATPIPFADGIVLGELEAVLPSHSESYQLPLAVAWDDAQPSALTQQLALGRIRQGR
RVGFLTD
GFAVEAMARGILRGLGDRSRITGRTGTLEFLGTERLDRLEVTDDLPVHWLSAEQS
NSSLIVG
DLAMIKLIRHIFPGIHPEVEMTRYLTRAGYDHTAPLLGEVAHTDSSGRRSTLIIV
QGAIRNQ
GDAWNWMLNNLRRGADELVLNDPAVQPGDDVFRPLISFVAMVGLRLGELHVVLAG
DTGDEAF
SPVVAGDSEVEAIKKAVAGEVAYALSKLDERAENADPAIDLLAAPLLERRSELVE
LAATLAE
STRDMLMTRTHGDFHLGQILVSEGDAVIIDFEGEPAKNLTERRAKTNPLRDVAGL
LRSLNYL
VATAQLDNDAVIEHENEVRREAIARFGRHAEEAFLDAYSQAVSVSKELDMPPDQR
RRVLDAF
LLEKAAYEIAYEARNRPKWLPIPLSGFTEIVSRLAGVKA

Rhizobium leguminosarum bv *viciae* UPM1137
MDTMNADSAPQPLWYKDAIIYQLHIKSFYDGNGDGVGDFAGLHQKLDHIAALGVN
AIWLLPF
FPSPRRDDGYDIADYGNVSPDYGTMEDFRAFVDAAHQRNIRVIIELVINHTSDQH
PWFQRAR
QAPAGSPERDFYVWSDTDQKFPETRIIFIDTEKSNWTWDAVAGAYYWHRFYSHQP
DLNFDSP
LVMEELLRVMRFWLETGIDGFRLDAIPYLVEREGTINENLPETHGILKRIRAALD
ATHPGVM
LLAEANQWPEDTREYFGEGDECHMAFHFPLMPRMYMAIAKEDRFPITDILRQTPE
IPDNCQW
AIFLRNHDELTLEMVTDAERDYLWETYASDKRARINLGIRRRLAPLMERDRRRIE
LMNALLL
SMPGTPVIYYGDEIGMGDNIYLGDRDGVRTPMQWSPDRNGGFSRADPARLVLPPV
ADPYGFE
AVNVEAQSTDAHSLLNWTRRMLALRGRHPAFGRGTLRFLSPENRKILAYLREYEG
EVLLCVA

SLSRLPQAVELDLSSFEGRVPIELTGMSPFPPIGQLTYLLTLPPYGFFWFQLTAD
ADPPAWR
TAPPEQLPDLLTMVIRRSLLDLVDEPGHARILSGEILPAYLSKRRWFGAKDQAVQ
AARLISA
TPIPFADGVVLGELEVELPNHSESYQLPLTVAWDDAHPSALAQQLALGRIRQGRR
VGFLTDG
FAVEAMARGILHGLRDRSRTTGRTGTLEFIGTERLDSLDISDELPVHWLSAEQSN
SSLLVGD
VAMIKLIRHIFPGIHPEVEMTRYLTRAGYDHTAPLLGEVAHTDSSGRRSTLIIVQ
GAIRNQG
DAWNWMLNNLRRGADELVLNDPAVQPDDDVFRSLISFVAMVGLRLGELHVVLAAK
TGDEAFS
PVVSGDSEVEAMRKAVAGELAYAMSKLEERDENADPAIDLLARPLLERRSELAEL
AATLAES
ARHTLMTRTHGDFHLGQILVSEGDAVIIDFEGEPAKNLAERRAKTNPLRDVAGLL
RSLSYLV
ATAQLDNDAVIEHDNEVRREAIARFGRHAEEAFLDAYSQAVSVSKELDMPPHQRR
RVLDAFL
LEKAAYEIAYEARNRPKWLPIPLFGLTEIVARLTGVKA

Rhizobium leguminosarum bv *trifolii* TA1
MDTMNADSAPLPLWYKDAIIYQLHIKSFYDANGDGVGDFAGLHQKLDHIAALGVN
AIWLLPF
FPSPRRDDGYDIADYGNVSPDYGTMEDFRAFVDAAHQRNIRVIIELVINHTSDQH
PWFQRAR
QAPAGSPERDFYVWSDTDQKFPETRIIFIDTEKSNWTWDAVAGAYYWHRFYSHQP
DLNFDSP
LVMEELLRVMRFWLTGIDGFRLDAIPYLVEREGTINENLPETHVILKRIRAALDA
THPGVML
LAEANQWPEDTREYFGEGDECHMAFHFPLMPRMYMAIAKEDRFPITDILRQTPEI
PDNCQWA
IFLRNHDELTLEMVTDAERDYLWETYASDKRARINLGIRRRLAPLMERDRRRIEL
MNALLLS
MPGTPVIYYGDEIGMGDNIYLGDRDGVRTPMQWSPDRNGGFSRADPARLVLPPVA
DPLYGFE
AVNVEAQSTDAHSLLNWTRRMLALRGRHPAFGRGTLRFLSPENRKILAYLREYEG
EVLLCVA
SLSRLPQAVELDLSSFEGRVPIELTGMSPFPPIGQLTYLLTLPPYGFFWFQLTAD
ADPPAWR
TAPPEQLPDLLTMVIRRNLLDLVDEPGHARILSGEILPAYLSKRRWFGAKDQALQ
AARLISA
TPIPFADGVVLGELEVVLPNHSESYQLPLAVAWDDAHPSALAQQLALGRIRQGRR
VGFLTDG
FAVEAMARGILHGLRDRSRTTGRTGTLEFIGTEQLDSLDISDELPVHWLSAEQSN
SSLLVGD
VAMIKLIRHIFPGIHPEVEMTRYLTRAGYDHTAPLLGEVAHTDSSGRRSTLIIVQ
GAIRNQG
DAWNWMLNNLRRGADELVLNDPAVQPDDDVFRSLISFVAMVGLRLGELHVVLAAK
TGDEAFS
PVVAGDSEVEAMRKAVAGELAYAMSKLDERDENADPAIDLLAAPLLKRRSELADL
AATLAES
ARHTLMTRTHGDFHLGQILVSEGDAVIIDFEGEPAKNLAERRAKTNPLRDVAGLL
RSLSYLV

ATAQLDNDAVIEHDNEVRREAIARFGRHAEEAFLDAYSQAVSVSKELDMPANQRR
RVLDAFL
LEKAAYEIAYEARNRPKWLPIPLSGLTEIVSRLAGVKA

Rhizobium leguminosarum bv *trifolii* CC278f
MDTMNADSAPQPLWYKDAIIYQLHIKSFYDANGDGVGDFAGLHQKLDHIAALGVN
AIWLLPF
FPSPRRDDGYDIADYGNVSPDYGTMEDFRAFVDAAHQRNIRVIIELVINHTSDQH
PWFQRAR
QAPAGSPERDFYVWSDTDQKFPETRIIFIDTEKSNWTWDAVAGAYYWHRFYSHQP
DLNFDSP
LVMEELLRVMRFWLETGIDGFRLDAIPYLVEREGTINENLPETHVILKRIRAALD
ATHPGVM
LLAEANQWPEDTREYFGEGDECHMAFHFPLMPRMYMAIAKEDRFPITDILRQTPE
IPDNCQW
AIFLRNHDELTLEMVTDAERDYLWETYASDKRARINLGIRRRLAPLMERDRRRIE
LMNALLL
SMPGTPVIYYGDEIGMGDNIYLGDRDGVRTPMQWSPDRNGGFSRADPARLVLPPV
ADPLYGF
EAVNVEAQSTDAHSLLNWTRRMLALRGRHPAFGRGTLRFLSPENRKILAYLREYE
GEVLLCV
ASLSRLPQAVELDLSSFEGRVPIELTGMSPFPPIGQLTYLLTLPPYGFFWFQLTA
DADPPAW
RTAPPEQLPDLLTMVIRRSLLDLVDEPGHARILSGEILPAYLSKRRWFGAKDQAL
QAARLIS
ATPIPFADGVVLGELEVVLPNHSESYQLPLAVAWDDAHPSALAQQLALGRIRQGR
RVGFLTD
GFAVEAMARGILHGLRDRSRTTGRTGTLEFIGTERLDSLDISDELPVHWLSAEQS
NSSLLVG
DVAMIKLIRHIFPGIHPEVEMTRYLTRAGYDHTAPLLGEVAHTDSSGRRSTLIIV
QGAIRNQ
GDAWNWMLNNLRRGADELVLNDPAVQPDDDVFQSLISFVAMVGLRLGELHVVLAA
KTGDEAF
SPVVSGDSEVEAMKKAVSGEVAYAMSKLDEREQNADPAIDLLAAPLLERRSELAE
LAEMLAE
SARHTLITRTHGDFHLGQILVSEGDAVIIDFEGEPAKNLTERRAKTNPLRDVAGL
LRSLSYL
VATAQLDNDAVIEHDNEVRREAIARFGRNAEEAFLDAYSQAVSVSKELDMPANQR
RRVLDAF
LLEKAAYEIAYEARNRPKWLPIPLSGFTEIVSRLTGVKA

Rhizobium leguminosarum bv *trifolii* CC278f
MDTMNADSAPQPLWYKDAIIYQLHIKSFYDANGDGVGDFAGLHQKLDHIAALGVN
AIWLLPF
FPSPRRDDGYDIADYGNVSPDYGTMEDFRAFVDAAHQRNIRVIIELVINHTSDQH
PWFQRAR
QAPAGSPERDFYVWSDTDQKFPETRIIFIDTEKSNWTWDAVAGAYYWHRFYSHQP
DLNFDSP
LVMEELLRVMRFWLETGIDGFRLDAIPYLVEREGTINENLPETHVILKRIRAALD
ATHPGVM
LLAEANQWPEDTREYFGEGDECHMAFHFPLMPRMYMAIAKEDRFPITDILRQTPE
IPDNCQW

AIFLRNHDELTLEMVTDAERDYLWETYASDKRARINLGIRRRLAPLMERDRRRIE
LMNALLL
SMPGTPVIYYGDEIGMGDNIYLGDRDGVRTPMQWSPDRNGGFSRADPARLVLPPV
ADPLYGF
EAVNVEAQSTDAHSLLNWTRRMLALRGRHPAFGRGTLRFLSPENRKILAYLREYE
GEVLLCV
ASLSRLPQAVELDLSSFEGRVPIELTGMSPFPPIGQLTYLLTLPPYGFFWFQLTA
DADPPAW
RTAPPEQLPDLLTMVIRRSLLDLVDEPGHARILSGEILPAYLSKRRWFGAKDQAL
QAARLIS
ATPIPFADGVVLGELEVVLPNHSESYQLPLAVAWDDAHPSALAQQLALGRIRQGR
RVGFLTD
GFAVEAMARGILHGLRDRSRTTGRTGTLEFIGTERLDSLDISDELPVHWLSAEQS
NSSLLVG
DVAMIKLIRHIFPGIHPEVEMTRYLTRAGYDHTAPLLGEVAHTDSSGRRSTLIIV
QGAIRNQ
GDAWNWMLNNLRRGADELVLNDPAVQPDDDVFQSLISFVAMVGLRLGELHVVLAA
KTGDEAF
SPVVSGDSEVEAMKKAVSGEVAYAMSKLDEREQNADPAIDLLAAPLLERRSELAE
LAEMLAE
SARHTLITRTHGDFHLGQILVSEGDAVIIDFEGEPAKNLTERRAKTNPLRDVAGL
LRSLSYL
VATAQLDNDAVIEHDNEVRREAIARFGRNAEEAFLDAYSQAVSVSKELDMPANQR
RRVLDAF
LLEKAAYEIAYEARNRPKWLPIPLSGFTEIVSRLTGVKA

A 2.3 *ots*B isolate nucleotide sequence

A 2.3.1 Isolate 32

TACTCAGGAAATGCTTTCACTGGTCTTGGAAGAGCCCGACCACTGGGCGATGTTT
CTCGATA
TCGACGGGACGCTGCTCAATCTTGCGCCCACGCCGGATGCGATCGAGGTCCCAGA
AGCGCTT
CCGGGACAACTTCATCGCCTGTCGAACAAGCTCGGCGGCGCCTTGGCGCTGGTCA
CCGGCCG
ATCACTTTCTTATGCAGACGAGCTGTTCAAACCCTTTGAATTTCCGACGGCAGGC
CTTCACG
GCGCCGAAATCAGGAACGCCGCAGGAATGCAGACGGTCGAGGCGACAGCTGAGTT
TCAGGCA
CTGAAGCATGCGCTGACCGCGGAAGCCGAGCACTATCCTGGCGTCTTGATCGAGG
ACAAGGG
CGCGGCCGTTGCCGCCCATTACCGGCTGGCGCCCGAATATGAAAAGGTGCTCGAA
GACCGCA
TGCACCATTATGCCGAGCTCGCCGGGCCGAATTGGGCATTGCAGCTCGGCAAGAT
GGTATTC
GAGCTGCGTCCGGCCCGGTCGAGCAAGGGCGATGCGCTGGAGCGGTTCTTCCAGT
CCGACCC
GTTCAAGAACCGCTGCCCGATTACCATCGGCGATGATCTGACCGACGAGTCGATG
TTTGCGA
TCGCCAATGCACGTGGTGGCGTTTCCGTGCGCGTCGGCGCGATCGGTGCTCCGAG
CTGCGCC
ACGAGCCGGCTGTCCTCTTCTGCGCTGGTCAGAAA

A 2.3.2 Isolate 40

TACTCAGGAAATGCTTTCGCTGGTCTTGGAAGAGCCCGATCACTGGGCGATGTTT
CTCGATA
TCGATGGGACGCTGCTCAATCTAGCGCCCACACCGGATGCGATCGAGGTCCCGGA
GGCGCTT
CCCGGACAACTTCATCGCCTGTCGAACAAGCTCGGCGGCGCCCTGGCACTGGTCA
CCGGACG
ATCGCTTGCCTATGCAGACGAGCTGTTCAAACCCTTTGCATTTCCGACAGCAGGT
CTGCACG
GCGCGGAAATCAGGAACGCCGCGGGAATGCAGACGGTCGAGGCGACACCTGAGTT
TCAGGCG
CTGAAGCATGCGCTGACCGCGGAAGCCGAGCATTATCCTGGAGTCTTGATCGAGG
ATAAGGG
TGCGGCCGTTGCCGCCCATTATCGGCTCGCGCCCGAATATGAAAAGGTGCTTGAA
GATCGCA
TGCGCCACTATGCGGAGGTCGCCGGGCCGAACTGGGCATTGCAGCTCGGCAAGAT
GGTATTC
GAACTGCGTCCGGCACGGTCGAGCAAGGGCGATGCGCTGGAGCGGTTTTTCCAGT
CCGACCC
CTTCAAGAACCGCTGCCCGATTACGATCGGCGACGATCTGACCGACGAGTCGATG
TTTGCGA
TCGCCAATGCGCGTGGTGGTGTTTCCGTGCGTGTCGGCGCGATCGGCACGCCAAG
CTGCGCC
ACGAGCCGACTGTCTTCTTCTGCGCTGGTCAGAAA

A 2.3.3 Isolate 41

TACTCAGGAAATGCTTTCGCTGGTCTTGGAAGAGCCCGATCACTGGGCGATGTTT
CTCGATA
TCGATGGGACGCTGCTCAATCTAGCGCCCACACCGGATGCGATCGAGGTCCCGGA
GGCGCTT
CCCGGACAACTTCATCGCCTGTCGAACAAGCTCGGCGGCGCCCTGGCACTGGTCA
CCGGACG
ATCGCTTGCCTATGCAGACGAGCTGTTCAAACCCTTTGCATTTCCGACAGCAGGT
CTGCACG
GCGCGGAAATCAGGAACGCCGCGGGAATGCAGACGGTCGAGGCGACACCTGAGTT
TCAGGCG
CTGAAGCATGCGCTGACCGCGGAAGCCGAGCATTATCCTGGAGTCTTGATCGAGG
ATAAGGG
TGCGGCCGTTGCCGCCCATTATCGGCTCGCGCCCGAATATGAAAAGGTGCTTGAA
GATCGCA
TGCGCCACTATGCGGAGGTCGCCGGGCCGAACTGGGCATTGCAGCTCGGCAAGAT
GGTATTC
GAACTGCGTCCGGCACGGTCGAGCAAGGGCGATGCGCTGGAGCGGTTTTTCCAGT
CCGACCC
CTTCAAGAACCGCTGCCCGATTACGATCGGCGACGATCTGACCGACGAGTCGATG
TTTGCGA
TCGCCAATGCGCGTGGTGGTGTTTCCGTGCGTGTCGGCGCGATCGGCACGCCAAG
CTGCGCC
ACGAGCCGACTGTCTTCTTCTGCGCTGGTCAGAAA

A 2.3.4 Isolate 42

CTACTCAGGAAATGCTTTCACTGGTCTTGGAAGAGCCCGACCACTGGGCGATGTT
TCTCGAT
ATCGACGGGACGCTGCTCAATCTTGCGCCCACGCCGGATGCGATCGAGGTCCCAG
AAGCGCT
TCCGGGACAACTTCATCGCCTGTCGAACAAGCTCGGCGGCGCCTTGGCGCTGGTC
ACCGGCC
GATCACTTTCTTATGCAGACGAGCTGTTCAAACCCTTTGAATTTCCGACGGCAGG
CCTTCAC
GGCGCCGAAATCAGGAACGCCGCAGGAATGCAGACGGTCGAGGCGACAGCTGAGT
TTCAGGC
ACTGAAGCATGCGCTGACCGCGGAAGCCGAGCACTATCCTGGCGTCTTGATCGAG
GACAAGG
GCGCGGCCGTTGCCGCCCATTACCGGCTGGCGCCCGAATATGAAAAGGTGCTCGA
AGACCGC
ATGCACCATTATGCCGAGCTCGCCGGGCCGAACTGGGCATTGCAGCTCGGCAAGA
TGGTATT
CGAGCTGCGTCCGGCCCGGTCGAGCAAGGGCGATGCGCTGGAGCGGTTCTTCCAG
TCCGACC
CGTTCAAGAACCGCTGCCCGATTACCATCGGCGATGATCTGACTGACGAGTCGAT
GTTTGCG
ATCGCCAATGCGCGTGGTGGCGTTTCCGTGCGCGTCGGCGCGATCGGTGCTCCGA
GCTGCGC
CACGAGCCGGCTGTCCTCTTCTGCGCTGGTCAGAA

A 2.3.5 Isolate 43

TACTCAGGAAATGCTTTCACTGGTCTTGGAAGAGCCCGATCACTGGGCGATGTTT
CTCGATA
TCGACGGGACGCTGCTCAATCTTGCGCCCACGCCGGATGCGATCGAGGTCCCAGA
GGCGCTT
CCCGGACAACTTCATCGCCTGTCGAACAAGCTCGGCGGTGCCCTGGCGCTGGTCA
CCGGACG
ATCGCTTGCCTATGCAGACGAGCTGTTCAAGCCCTTTGCATTTCCGACAGCAGGC
CTGCACG
GCGCGGAAATCAGGAACGCCGCTGGAATGCAGACGATCGAGGCGACACCTGAATT
TCAGGCG
CTGAAGCATGCGCTGACCGCAGAAGCCGAGCACTACCCTGGCGTCTTAATCGAGG
ACAAGGG
CGCGGCCGTTGCCGCCCATTACCGGCTCGCACCCGAATATGAAAAGGTACTCGAA
GAGCGCA
TGCACCACTATGCGGAGATCGCCGGGCCGAACTGGGCATTGCAGCTTGGCAAGAT
GGTGTTC
GAGCTGCGTCCGGCACGATCGAGCAAGGGCGATGCGCTGGAGCGGTTTTTCCAGT
CCGACCC
GTTCAAGAACCGCTCCCCGATTACCATCGGCGATGATCTGACCGACGAGTCGATG
TTTGCAA
TCGCCAATGCCCGCGGCGGCCTTTCCGTGCGTGTCGGCGCGATCGGCGCCCCGAG
CTGCGCC
ACGAGCACGGCTGTCTTCTTCTGCGCTGGTCAGA

A 2.3.6 Isolate 47

TACTCAGGAAATGCTTTCACTGGTCTTGGAAGAGCCCGATCACTGGGCGATGTTT
CTCGATA
TCGACGGGACGCTGCTCAATCTTGCGCCCACGCCGGATGCGATCGAGGTCCCAGA
GGCGCTT
CCCGGACAACTTCATCGCCTGTCGAACAAGCTCGGCGGTGCCCTGGCGCTGGTCA
CCGGACG
ATCGCTTGCCTATGCAGACGAGCTGTTCAAGCCCTTTGCATTTCCGACAGCAGGC
CTGCACG
GCGCGGAAATCAGGAACGCCGCTGGAATGCAGACGATCGAGGCGACACCTGAATT
TCAGGCG
CTGAAGCATGCGCTGACCGCAGAAGCCGAGCACTACCCTGGCGTCTTAATCGAGG
ACAAGGG
CGCGGCCGTTGCCGCCCATTACCGGCTCGCACCCGAATATGAAAAGGTACTCGAA
GAGCGCA
TGCACCACTATGCGGAGATCGCCGGGCCGAACTGGGCATTGCAGCTTGGCAAGAT
GGTGTTC
GAGCTGCGTCCGGCACGATCGAGCAAGGGCGATGCGCTGGAGCGGTTTTTCCAGT
CCGACCC
GTTCAAGAACCGCTCCCCGATTACCATCGGCGATGATCTGACCGACGAGTCGATG
TTTGCAA
TCGCCAATGCCCGCGGCGGCCTTTCCGTGCGTGTCGGCGCGATCGGCGCCCCGAG
CTGCGCC
ACGAGCCGGCTGTCTTCTTCTGCGCTGGTCAGA

A 2.3.7 Isolate 50

TACTCAGGAAATGCTTTCACTGGTCTTGGAAGAGCCCGATCACTGGGCGATGTTT
CTCGATA
TCGACGGGACGCTGCTCAATCTTGCGCCCACGCCGGATGCGATCGAGGTCCCAGA
GGCGCTT
CCCGGACAACTTCATCGCCTGTCGAACAAGCTCGGCGGTGCCCTGGCGCTGGTCA
CCGGACG
ATCGCTTGCCTATGCAGACGAGCTGTTCAAGCCCTTTGCATTTCCGACAGCAGGC
CTGCACG
GCGCGGAAATCAGGAACGCCGCTGGAATGCAGACGATCGAGGCGACACCTGAATT
TCAGGCG
CTGAAGCATGCGCTGACCGCAGAAGCCGAGCACTACCCTGGCGTCTTAATCGAGG
ACAAGGG
CGCGGCCGTTGCCGCCCATTACCGGCTCGCACCCGAATATGAAAAGGTACTCGAA
GAGCGCA
TGCACCACTATGCGGAGATCGCCGGGCCGAACTGGGCATTGCAGCTTGGCAAGAT
GGTGTTC
GAGCTGCGTCCGGCACGATCGAGCAAGGGCGATGCGCTGGAGCGGTTTTTCCAGT
CCGACCC
GTTCAAGAACCGCTCCCCGATTACCATCGGCGATGATCTGACCGACGAGTCGATG
TTTGCAA
TCACCAATGCCCGCGGCGGCCTTTCCGTGCGTGTCGGCGCGATCGGCGCCCCTAG
CTGCGCC
ACGAGCCGGCTGTCTTCTTCTGCGCTGGTCAGA

A 2.3.8 Isolate 51

TACTCAGGAAATGCTTTCACTGGTCTTGGAAGAGCCCGATCACTGGGCGATGTTT
CTCGATA
TCGACGGGACGCTGCTCAATCTTGCGCCCACGCCGGATGCGATCGAGGTCCCAGA
GGCGCTT
CCCGGACAACTTCATCGCCTGTCGAACAAGCTCGGCGGTGCCCTGGCGCTGGTCA
CCGGACG
ATCGCTTGCCTATGCAGACGAGCTGTTCAAGCCCTTTGCATTTCCGACAGCAGGC
CTGCACG
GCGCGGAAATCAGGAACGCCGCTGGAATGCAGACGATCGAGGCGACACCTGAATT
TCAGGCG
CTGAAGCATGCGCTGACCGCAGAAGCCGAGCACTACCCTGGCGTCTTAATCGAGG
ACAAGGG
CGCGGCCGTTGCCGCCCATTACCGGCTCGCACCCGAATATGAAAAGGTACTCGAA
GAGCGCA
TGCACCACTATGCGGAGATCGCCGGGCCGAACTGGGCATTGCAGCTTGGCAAGAT
GGTGTTC
GAGCTGCGTCCGGCACGATCGAGCAAGGGCGATGCGCTGGAGCGGTTTTTCCAGT
CCGACCC
GTTCAAGAACCGCTCCCCGATTACCATCGGCGATGATCTGACCGACGAGTCGATG
TTTGCAA
TCACCAATGCCCGCGGCGGCCTTTCCGTGCGTGTCGGCGCGATCGGCGCCCCTAG
CTGCGCC
ACGAGCCGGCTGTCTTCTTCTGCGCTGGTCAGAAA

A 2.3.9 Isolate 53

CTACTCAGGAAATGCTTTCGCTGGTCTTGGAAGAGCCCGATCACTGGGCGATGTT
TCTCGAT
ATCGATGGGACGCTGCTCAATCTAGCGCCCACACCGGATGCGATCGAGGTCCCGG
AGGCGCT
TCCCGGACAACTTCATCGCCTGTCGAACAAGCTCGGCGGCGCCCTGGCACTGGTC
ACCGGAC
GATCGCTTGCCTATGCAGACGAGCTGTTCAAACCCTTTGCATTTCCGACAGCAGG
TCTGCAC
GGCGCGGAAATCAGGAACGCCGCGGGAATGCAGACGGTCGAGGCGACACCTGAGT
TTCAGGC
GCTGAAGCATGCGCTGACCGCGGAAGCCGAGCATTATCCCGGAGTCTTGATCGAG
GATAAGG
GTGCGGCCGTTGCCGCCCATTATCGGCTCGCGCCCGAATATGAAAAAGTGCTTGA
AGATCGC
ATGCGCCACTATGCGGAGGTCGCCGGGCCGAACTGGGCATTGCAGCTCGGCAAGA
TGGTATT
CGAACTGCGTCCGGCACGGTCGAGCAAGGGCGATGCGCTGGAGCGGTTTTTCCAG
TCCGACC
CCTTCAAGAACCGCTGCCCGATTACGATCGGCGACGATCTGACCGACGAGTCGAT
GTTTGCG
ATCGCCAATGCGCGTGGTGGTGTTTCCGTGCGTGTCGGCGCGATCGGCACGCCAA
GCTGCGC
CACGAGCCGACTGTCTTCTTCTGCGCTGGT

A 2.3.10 TA1

GACGCTGCTCAATCTTGCGCCCACGCCGGATGCGATCGAGGTCCCAGAAGCGCTT
CCGGGAC
AACTTCATCGCCTGTCGAACAAGCTCGGCGGCGCCTTGGCGCTGGTCACCGGCCG
ATCACTT
TCTTATGCAGACGAGCTGTTCAAACCCTTTGAATTTCCGACGGCAGGCCTTCACG
GCGCCGA
AATCAGGAACGCCGCAGGAATGCAGACGGTCGAGGCGACAGCTGAGTTTCAGGCA
CTGAAGC
ATGCGCTGACCGCGGAAGCCGAGCACTATCCTGGCGTCTTGATCGAGGACAAGGG
CGCGGCC
GTTGCCGCCCATTACCGGCTGGCGCCCGAATATGAAAAGGTGCTCGAAGACCGCA
TGCACCA
TTATGCCGAGCTCGCCGGGCCGAACTGGGCATTGCAGCTCGGCAAGATGGTATTC
GAGCTGC
GTCCGGCCCGGTCGAGCAAGGGCGATGCGCTGGAGCGGTTCTTCCAGTCCGACCC
GTTCAAG
AACCGCTGCCCGATTACCATCGGCGATGATCTGACTGACGAGTCGATGTTTGCGA
TCGCCAA
TGCGCGTGGTGGCGTTTCCGTGCGCGTCGGCGCGATCGGTGCTCCGAGCTGCGCC
ACGAGCC
GGCTGTCCTCTTCT
GCGCTG

A 2.3.11 WSM1325

ATCTTGCACCCACACCGGATGCGATAGAGGTTCCAGAGGCGCTTCCTGGACAACT
TCATCGC
CTGTCGAACAAGCTCGGCGGCGCCTTGGCGCTGGTCACCGGACGGTCTCTTGCTT
ATGCAGA
CGCGCTGTTCAAACCCTTTGCATTTCCGACAGCAGGCCTGCACGGCGCCGAAATC
AGGAACG
CCGCCGGAATGCAGACGGTCGAGGCGACACCTGAGTTTCAGGCGTTGAAACATGC
GCTGACC
GCTGAAGCCGAGCACTACCCTGGCGTCCTGATCGAGGACAAGGGCGCGGCCGTTG
CCGCCCA
CTACCGGCTGGCGCCGGAATATGAAAAGGTGCTCGAAGACCGCATGCACCACTAT
GCGGAGC
TCGCCGGGCCGAACTGGGCATTGCAGCTCGGCAAGATGGTATTCGAGCTGCGCCC
GGCACGG
TCGAGCAAGGGCGATGCGCTGGAGCGGTTCTTCCAGTCCGACCCGTTCAAGAACC
GCTCGCC
GATTACCATTGGCGATGATCTGACCGACGAGTCGATGTTTGCGATCGCCAATGCG
CGTGGTG
GCGTTTCCGTGCGTGTCGGCGCGATCGGCGCTCCCAGCTGCGCCACGAGCCGGCA
TGTCCTC
TTC

A 2.3.12 cc275e

GGGACGCTGCTCAATCTAGCGCCCACACCGGATGCGATCGAGGTCCCGGAGGCGC
TTCCCGG
ACAACTTCATCGCCTGTCGAACAAGCTCGGCGGCGCCCTGGCACTGGTCACCGGA
CGATCGC

TTGCCTATGCAGACGAGCTGTTCAAACCCTTTGCATTTCCGACAGCAGGTCTGCA
CGGCGCG
GAAATCAGGAACGCCGCGGGAATGCAGACGGTCGAGGCGACACCTGAGTTTCAGG
CGCTGAA
GCATGCGCTGACCGCGGAAGCCGAGCATTATCCCGGAGTCTTGATCGAGGATAAG
GGTGCGG
CCGTTGCCGCCCATTATCGGCTCGCGCCCGAATATGAAAAAGTGCTTGAAGATCG
CATGCGC
CACTATGCGGAGGTCGCCGGGCCGAACTGGGCATTGCAGCTCGGCAAGATGGTAT
TCGAACT
GCGTCCGGCACGGTCGAGCAAGGGCGATGCGCTGGAGCGGTTTTTCCAGTCCGAC
CCCTTCA
AGAACCGCTGCCCGATTACGATCGGCGACGATCTGACCGACGAGTCGATGTTTGC
GATCGCC
AATGCGCGTGGTGGTGTTTCCGTGCGTGTCGGCGCGATCGGCACGCCAAGCTGCG
CCACGAG
CCGACTG

A 2.4 Isolate *ots*B amino acid sequence

A 2.4.1 Isolate 32

TQEMLSLVLEEPDHWAMFLDIDGTLLNLAPTPDAIEVPEALPGQLHRLSNKLGGA
LALVTGR
SLSYADELFKPFEFPTAGLHGAEIRNAAGMQTVEATAEFQALKHALTAEAEHYPG
VLIEDKG
AAVAAHYRLAPEYEKVLEDRMHHYAELAGPNWALQLGKMVFELRPARSSKGDALE
RFFQSDP
FKNRCPITIGDDLTDESMFAIANARGGVSVRVGAIGAPSCATSRLSSSALVR

A 2.4.2 Isolate 40

TQEMLSLVLEEPDHWAMFLDIDGTLLNLAPTPDAIEVPEALPGQLHRLSNKLGGA
LALVTGR
SLAYADELFKPFAFPTAGLHGAEIRNAAGMQTVEATPEFQALKHALTAEAEHYPG
VLIEDKG
AAVAAHYRLAPEYEKVLEDRMRHYAEVAGPNWALQLGKMVFELRPARSSKGDALE
RFFQSDP
FKNRCPITIGDDLTDESMFAIANARGGVSVRVGAIGTPSCATSRLSSSALVR

A 2.4.3 Isolate 41

TQEMLSLVLEEPDHWAMFLDIDGTLLNLAPTPDAIEVPEALPGQLHRLSNKLGGA
LALVTGR
SLAYADELFKPFAFPTAGLHGAEIRNAAGMQTVEATPEFQALKHALTAEAEHYPG
VLIEDKG
AAVAAHYRLAPEYEKVLEDRMRHYAEVAGPNWALQLGKMVFELRPARSSKGDALE
RFFQSDP
FKNRCPITIGDDLTDESMFAIANARGGVSVRVGAIGTPSCATSRLS
SSALVR

A 2.4.4 Isolate 42

TQEMLSLVLEEPDHWAMFLDIDGTLLNLAPTPDAIEVPEALPGQLHRLSNKLGGA
LALVTGR
SLSYADELFKPFEFPTAGLHGAEIRNAAGMQTVEATAEFQALKHALTAEAEHYPG
VLIEDKG
AAVAAHYRLAPEYEKVLEDRMHHYAELAGPNWALQLGKMVFELRPARSSKGDALE
RFFQSDP
FKNRCPITIGDDLTDESMFAIANARGGVSVRVGAIGAPSCATSRLSSSALVR

A 2.4.5 Isolate 43

TQEMLSLVLEEPDHWAMFLDIDGTLLNLAPTPDAIEVPEALPGQLHRLSNKLGGA
LALVTGR
SLAYADELFKPFAFPTAGLHGAEIRNAAGMQTIEATPEFQALKHALTAEAEHYPG
VLIEDKG
AAVAAHYRLAPEYEKVLEERMHHYAEIAGPNWALQLGKMVFELRPARSSKGDALE
RFFQSDP
FKNRSPITIGDDLTDESMFAIANARGGLSVRVGAIGAPSCATSTAVFFCAGQ

A 2.4.6 Isolate 47

TQEMLSLVLEEPDHWAMFLDIDGTLLNLAPTPDAIEVPEALPGQLHRLSNKLGGA
LALVTGR
SLAYADELFKPFAFPTAGLHGAEIRNAAGMQTIEATPEFQALKHALTAEAEHYPG
VLIEDKG
AAVAAHYRLAPEYEKVLEERMHHYAEIAGPNWALQLGKMVFELRPARSSKGDALE
RFFQSDP
FKNRSPITIGDDLTDESMFAIANARGGLSVRVGAIGAPSCATSRLSSSALVR

A 2.4.7 Isolate 50

TQEMLSLVLEEPDHWAMFLDIDGTLLNLAPTPDAIEVPEALPGQLHRLSNKLGGA
LALVTGR
SLAYADELFKPFAFPTAGLHGAEIRNAAGMQTIEATPEFQALKHALTAEAEHYPG
VLIEDKG
AAVAAHYRLAPEYEKVLEERMHHYAEIAGPNWALQLGKMVFELRPARSSKGDALE
RFFQSDP
FKNRSPITIGDDLTDESMFAITNARGGLSVRVGAIGAPSCATSRLSSSALVR

A 2.4.8 Isolate 51

TQEMLSLVLEEPDHWAMFLDIDGTLLNLAPTPDAIEVPEALPGQLHRLSNKLGGA
LALVTGR
SLAYADELFKPFAFPTAGLHGAEIRNAAGMQTIEATPEFQALKHALTAEAEHYPG
VLIEDKG
AAVAAHYRLAPEYEKVLEERMHHYAEIAGPNWALQLGKMVFELRPARSSKGDALE
RFFQSDP
FKNRSPITIGDDLTDESMFAITNARGGLSVRVGAIGAPSCATSRLSSSALVR

A 2.4.9 Isolate 53

TQEMLSLVLEEPDHWAMFLDIDGTLLNLAPTPDAIEVPEALPGQLHRLSNKLGGA
LALVTGR
SLAYADELFKPFAFPTAGLHGAEIRNAAGMQTVEATPEFQALKHALTAEAEHYPG
VLIEDKG
AAVAAHYRLAPEYEKVLEDRMRHYAEVAGPNWALQLGKMVFELRPARSSKGDALE
RFFQSDP
FKNRCPITIGDDLTDESMFAIANARGGVSVRVGAIGTPSCATSRLSSSAL

A 2.4.10 TA1

TLLNLAPTPDAIEVPEALPGQLHRLSNKLGGALALVTGRSLSYADELFKPFEFPT
AGLHGAE
IRNAAGMQTVEATAEFQALKHALTAEAEHYPGVLIEDKGAAVAAHYRLAPEYEKV
LEDRMHH
YAELAGPNWALQLGKMVFELRPARSSKGDALERFFQSDPFKNRCPITIGDDLTDE
SMFAIAN
ARGGVSVRVGAIGAPSCATSRLSSSAL

A 2.4.11 WSM1325

LAPTPDAIEVPEALPGQLHRLSNKLGGALALVTGRSLAYADALFKPFAFPTAGLH
GAEIRNA
AGMQTVEATPEFQALKHALTAEAEHYPGVLIEDKGAAVAAHYRLAPEYEKVLEDR
MHHYAEL
AGPNWALQLGKMVFELRPARSSKGDALERFFQSDPFKNRSPITIGDDLTDESMFA
IANARGG
VSVRVGAIGAPSCAT
SRHVLF

A 2.4.12 CC275e

GTLLNLAPTPDAIEVPEALPGQLHRLSNKLGGALALVTGRSLAYADELFKPFAFP
TAGLHGA
EIRNAAGMQTVEATPEFQALKHALTAEAEHYPGVLIEDKGAAVAAHYRLAPEYEK
VLEDRMR
HYAEVAGPNWALQLGKMVFELRPARSSKGDALERFFQSDPFKNRCPITIGDDLTD
ESMFAIA
NARGGVSVRVGAIGTPSCATSRL